Sensitivity and Uncertainty Analysis of Episodic Ozone Predictions from the Community Multiscale Air Quality Model

Beddows, Andrew Victor

Awarding institution:
King's College London

Sensitivity and Uncertainty Analysis of Episodic Ozone Predictions from the Community Multiscale Air Quality Model

Andrew Beddows

A thesis submitted for the degree of

Doctor of Philosophy

in the

Analytical and Environmental Sciences Division

Faculty of Life Sciences and Medicine

King's College London

Abstract

The first global variance-based sensitivity analysis of ozone and NO_2 concentrations produced by the CMAQ model during the July 2006 ozone pollution episode over the UK has been performed. Gaussian process emulation methods have been employed to overcome the problems caused by long model run times which have previously prevented such analyses being undertaken. The computationally efficient Morris' method was used to rank the effect of perturbations in 223 model input variables, including all of the gas-phase species in the model domain chemical boundary conditions and emissions, and all of the reaction rates in the carbon bond five core chemical mechanism. The 30 most influential variables were combined with ozone deposition velocity to emulate the effects of perturbations in 31 input variables on the modelled concentrations of ozone and NO_2. These emulators were then used in place of CMAQ in Fourier amplitude sensitivity tests, which decompose the variance induced in model output when all of the inputs are perturbed together into contributions from each of those inputs. These tests were performed for every hour of a 21 day time series spanning the episode and several days either side, for a number of locations around the UK. The results reveal a complex spatio-temporal pattern of model sensitivities, with NO and isoprene emissions, NO_2 photolysis and ozone deposition velocity and boundary conditions being amongst the most influential input uncertainties. The same emulators were used in a Monte Carlo uncertainty analysis of modelled ozone concentrations. The results of this analysis were used with a simple Bayesian weighting procedure to calibrate the model inputs, which led to a significant improvement in peak afternoon ozone predictions. Calibrated UK and EU NO and NO_2 emissions were between 1.27 and 1.38 times the baseline values, suggesting that official NO_x emissions totals may be substantially underestimated.

Acknowledgements

The author would like to thank the following:

Dr. Sean Beevers of King's College London (KCL) for acting as first supervisor and providing appropriate computing facilities and advice on various aspects of air quality modelling.

Dr. Marta Blangiardo of Imperial College London for acting as second supervisor and providing advice on various aspects of statistics.

Dr. Nutthida Kitwiroon (KCL) for providing emissions and meteorological data for input to the CMAQ model.

Prof. Martin Williams (KCL) and Dr. David Carslaw (formerly KCL) for providing advice and feedback throughout the research.

Contents

List of Figures

List of Tables

Abreviations

AQMEII	Air Quality Model Evaluation International Initiative
BC(s)	Boundary condition(s)
BCON	Boundary conditions processor
BMC	Bayesian Monte Carlo
CB-IV	Carbon bond four chemical mechanism
CB-05	Carbon bond five chemical mechanism
CMAQ	Community multiscale air quality (model)
CMAS	Community modelling and analysis system
CTM	Chemistry transport model
CCTM	CMAQ chemistry transport model
DDM	Decoupled direct method
DEFRA	Department for the environment, food and rural affairs
EMEP	European monitoring and evaluation programme
FAST	Fourier amplitude sensitivity test
GLUE	Generalised likelihood uncertainty estimation
HDMR	High dimensional model representation
IC(s)	Initial condition(s)
ICON	Initial conditions processor
JPROC	Photolysis rate processor
LHS	Latin Hypercube sample
MAE	Mean absolute error
MB	Mean bias
MCIP	Meteorology - chemistry interface processor
NAEI	National atmospheric emissions inventory
netCDF	Network common data format
NO_x	Total of Nitric oxide and Nitrogen dioxide
PM	Particulate matter
PM_x	Particulate matter less than x microns in diameter
REVIHAAP	Review of evidence on health aspects of air pollution
RFM	Reduced form model

SMOKE	Sparse matrix operator kernel emissions
SAPRC-99	Statewide air pollution research centre 1999 chemical mechanism
SAPRC-07	Statewide air pollution research centre 2007 chemical mechanism
SOA	Secondary organic aerosol
USEPA	United States environmental protection agency
VOC	Volatile organic compound
WRF	Weather research and forecasting (model)

Chapter 1

Introduction

This thesis describes the sensitivity and uncertainty analysis of the Community Multiscale Air Quality (CMAQ) model for an ozone pollution episode during July 2006. This focus on a single episode has enabled advanced techniques not previously employed in air quality modelling to be used to great effect, and thus indicates the future direction that this kind of work should take. CMAQ is a Eulerian, or grid based, chemistry transport model (CTM) which is designed to predict or reproduce the concentrations of a number of chemical species over a wide range of spatial scales (Byun and Schere, 2006). It simulates both gas-phase and particulate matter (PM) chemistry and is in a process of continual development, with regular updates being released. This work concerns the gas-phase chemistry in version 4.7.1 of the model, focussing in particular on the simulated concentrations of ozone (O_3) and nitrogen dioxide (NO_2). Several new releases of the CMAQ code have taken place during the course of this project, with the major changes being in the way PM is represented. The gas-phase chemistry, however, has remained largely unchanged, meaning that the results presented here are still relevant.

This introductory chapter is organized as follows: Section 1.1 gives the motivation for modelling air quality from a health perspective, while section 1.2 describes tropospheric ozone chemistry and the related public policy implications are discussed in section 1.3. Section 1.4 contains a description of the CMAQ

modelling system, and section 1.5 introduces some basic concepts of sensitivity and uncertainty analysis. Following this, sections 1.6 and 1.7 describe the work of other authors in the area of sensitivity analysis and uncertainty analysis, focussing on CMAQ and similar models where possible, but by necessity extending the discussion to climate modelling, as more such work has been done in that area.

1.1 Air Pollution and Health

Whilst the protective effect of stratospheric ozone against the harmful effects of short-wave solar radiation is well known, at lower levels in the atmosphere it can be a harmful pollutant. Ozone exists naturally in the troposphere, but concentrations elevated above normal background levels are detrimental to human health and cause ecosystem wide damage (Madden and Hogsett, 2001). An up to date synthesis of evidence on the detrimental health effects of poor air quality can be found in the form of the World Health Organisation Review of Evidence on Health Aspects of Air Pollution (REVIHAAP, 2013), which draws together information on a range of pollutants including ozone and NO_2. Some of the references cited therein will be briefly reviewed here in order to highlight the importance of air quality modelling, and hence the motivation for this project.

Adverse health effects can be split into those caused by long-term and short-term exposure. In detecting long-term effects, the REVIHAAP report highlights difficulties caused by the lack of availability of long-term exposure data, and correlation of ozone levels with other pollutants, such as $PM_{2.5}$ (particulate matter with diameter less than 2.5 μ m). This was demonstrated in a large US cohort study by Jerrett et al. (2009), which followed nearly half a million subjects for 18 years. The authors were able to show an increased risk of death from cardiopulmonary causes for either ozone or $PM_{2.5}$, but when their statistical model included both pollutants, cardiovascular deaths were associated with $PM_{2.5}$ and deaths from respiratory causes with ozone. Smith et al. (2009) examined the same

data, this time separating out black carbon and sulfate as well as total $PM_{2.5}$ and reached similar conclusions. They point out, however, that this confounding by different pollutants in statistical models may be complicated by the toxicological effects of pollutant mixtures. These studies both use a large cohort with a long follow-up period, and both control for 20 potentially confounding factors, such as cigarette smoking, but Jerrett et al. (2009) note that some other studies have failed to find a significant relationship between ozone and mortality, highlighting the aforementioned difficulties associated with long-term exposure studies. In an investigation of the effects of elevated ozone levels on members of society who might be more susceptible, Zanobetti and Schwartz (2011) used cohorts of people who had been discharged from hospital after treatment for heart disease, chronic lung disease and diabetes. Survival data on such people from 105 US cities were used along with the yearly deviation of ozone concentration from the long-term trend for each particular city. The authors found a significant increased risk of mortality in all groups, but do point out, however, that they had no way of controlling for $PM_{2.5}$ exposure.

The studies discussed so far all concern the effects of ozone exposure on mortality rates, but evidence also exists for adverse effects of long-term exposure on respiratory morbidity. The focus of much of this work has been on asthma, with a review by Tzivian (2011) concluding that childhood hospital admissions may be increased by chronic exposure to ambient ozone. This finding is supported in a review by the United States Environmental Protection Agency (USEPA, 2011), which asserts that a number of epidemiological studies demonstrate associations between long-term exposure and increased symptoms in asthmatics, and new cases of asthma in children. Long-term ozone exposure has also been associated with decreased lung function growth, demonstrated in a three year study involving over three thousand children in Mexico City (Rojas-Martinez et al., 2007). Other possible effects of long-term exposure described in the REVIHAAP report include increased risk of preterm birth and adult cognitive decline.

A great deal of work has also been published on the short-term effects of ozone exposure, a good example being the APHENA study (Katsouyanni et al., 2009), which used data on hospital admissions from 134 cities across the US, Canada and Europe. The authors found positive associations between short-term exposure and both hospital admissions and mortality for cardiovascular and respiratory causes. Another study worthy of mention is the meta-analysis of world-wide English language literature by Ji et al. (2011), which reported positive associations between increased ozone exposure and hospital admissions for asthma, chronic obstructive pulmonary disease (COPD) and all respiratory diseases. Ozone pollution episodes the UK, which would give rise to such short-term exposures, tend to be associated with anticyclonic conditions and temperatures approaching 30 °C or above (Lee et al., 2006). In a study of causes of mortality during a pollution episode in the August 2003 heatwave, Stedman (2004) reported that between 225 and 593 deaths in the UK could be attributable to elevated ozone levels.

The epidemiological evidence reviewed here indicates that ozone has adverse effects on human health. There does exist a body of literature concerning studies which have failed to show any positive association between exposure and mortality or morbidity, but the general scientific consensus seems to be that there are 'known adverse effects' (REVIHAAP, 2013). The report goes on to detail some of the toxic effects of ozone which have been demonstrated in experimental studies, and also points out that such work may actually underestimate the toxicity of the mixture of of compounds in ozone-polluted air as a whole.

Moving on to evidence for the health effects of NO_2, again only a small fraction off the available literature will be cited, the aim being just to highlight a few of the most important studies which motivate the current work. Starting with the effects of short-term exposure on mortality, the World Health Organisation adopts the view that daily NO_2 levels are associated with increases in all-cause mortality and cardiovascular and respiratory mortality (REVIHAAP, 2013). There is, however, some debate about whether NO_2 is the sole cause of this increased

mortality due to its high correlation with other traffic related pollutants such as PM (CARB, 2007). A review and meta-analysis by Anderson et al. (2007) found, conversely, that effects estimates for mortality and hospital admissions were not greatly modified after adjustments for other pollutants. Elevated NO_2 concentrations have been associated with increased hospital admissions for asthma by the USEPA (2008), particularly amongst children and the elderly. This is supported by a study of hospital asthma admissions in relation to PM_{10}, in which the increase in admissions of 0-14 year olds, for a 10 μg m^{-3} concentration increase, fell from 1.2 to 0.1 % after the statistical model was adjusted for NO_2 concentrations (Atkinson et al., 2001).

A great deal of work on the effects of long-term exposure to NO_2 has also been undertaken, and once again much of this is summarized in some key reviews. CARB (2007) states that exposure of a year or more can lead to preterm birth, and in children can give rise to changes in lung function growth and asthmatic symptoms. A different review of epidemiological and toxicological evidence concluded that the evidence for increased mortality and morbidity in relation to long-term exposure was 'suggestive, but not sufficient to infer a causal relationship' (USEPA, 2008). REVIHAAP (2013) takes the overall view that NO_2 may have a direct effect on health, or may be acting as a marker for PM or for other gaseous pollutants.

Taken as a whole, for both ozone and NO_2, the evidence indicates general acceptance that both pollutants are harmful to human health, although there is uncertainty concerning the exact mechanisms of this detriment, and its severity for any given level of exposure. What is less uncertain is that the whole mixture of substances in polluted air is harmful to health, and that these two gases are an important part of that mixture. CMAQ attempts to model the complex chemistry of this pollutant mixture as it evolves through time over regional scales, so sensitivity analyses of the model output of ozone and NO_2 should provide valuable insights into regional scale air quality.

1.2 Tropospheric Ozone Chemistry

Ozone is a secondary pollutant, not being emitted directly into the atmosphere as a result of anthropogenic activity, but rather produced by reactions involving volatile organic compounds (VOC's) and NO_x (NO and NO_2) in the presence of sunlight (Jacob, 1999). The chemistry of NO_x and ozone is intimately linked in the following set of three reactions:

$$NO_2 + h\nu \longrightarrow NO + O \quad (\lambda < 430\,nm) \tag{1.1}$$

$$O + O_2 \xrightarrow{M} O_3 \tag{1.2}$$

$$NO + O_3 \longrightarrow NO_2 + O_2 \tag{1.3}$$

The photolysis in reaction, 1.1, is shown in a common form, with the incident photon represented as $h\nu$, h being Planck's constant and ν the photon frequency, so that these terms together represent the energy carried by the photon. This energy will be sufficient to initiate the reaction if the photon wavelength, λ, is less than 430 nm. These reactions form a cycle in which the photolysis of NO_2 is followed by the reaction of O with O_2, in the presence of another molecule, to form ozone. In the third reaction NO_2 is created and ozone is destroyed, so that in the absence of competing interconversion reactions there is no net production of ozone. This is a photostationary state relationship in which the concentrations of the three species are related by the expression (Jenkin and Clemitshaw, 2000),

$$[O_3] = \frac{j_1[NO_2]}{k_3[NO]} \tag{1.4}$$

where j_1 is the NO_2 photolysis rate, k_3 is the rate coefficient for reaction 1.3 and square brackets denote concentration. Reactions which convert NO to NO_2, but without consuming ozone as reaction 1.3 does, will cause a deviation from this photostationary state and lead to net ozone production. Two of the most important such reactions are,

$$HO_2 + NO \longrightarrow OH + NO_2 \tag{1.5}$$

$$RO_2 + NO \longrightarrow RO + NO_2 \tag{1.6}$$

where in both cases NO is oxidised by a peroxy radical, the hydroperoxy radical (HO_2) in equation 1.5, and an organic peroxy radical (RO_2) in equation 1.6. Equation 1.5 is also an important source of hydroxyl radical (OH), which is a major driver of daytime chemistry in both polluted and clean parts of the atmosphere (Finlayson-Pitts and Pitts, 2000). Anthropogenic and biogenic VOCs (represented below by RH) provide the source of RO_2 via an oxidation process which is initiated by OH:

$$RH + OH \longrightarrow R + H_2O \tag{1.7}$$

$$R + O_2 \xrightarrow{M} RO_2 \tag{1.8}$$

NO_2 then forms via equation 1.6, and ozone is formed during the process described by equations 1.1 and 1.2. As an example, if the VOC was ethane (C_2H_6), this would be the RH in equation 1.7, and R would be the ethyl radical (C_2H_5) which is oxidized in equation 1.8, to form RO_2, which in this case would be the ethylperoxy radical ($C_2H_5O_2$) (Jacobson, 2002). The production of NO_2 in equation 1.6 reduces RO_2 to RO, an alkoxy radical, which in our example

would be the ethoxy radical (C_2H_5O). Jenkin and Clemitshaw (2000) go on to describe how the alkoxy radical is broken down by further oxidation or thermal decomposition, into carbonyl products and HO_2, which then creates further NO_2 and regenerates OH by reaction 1.5. The exact number of steps in this whole VOC breakdown process depends on the size and complexity of the particular VOC species, and may result in the creation of several ozone molecules for each VOC molecule.

A further source of OH is ozone photolysis, producing $O(^1D)$, which reacts with water vapour to form two OH molecules. Therefore, by the above processes, a single ozone photolysis may result in the creation of a number of new ozone molecules, hence its description by Jenkin and Clemitshaw as an autocatalyst.

This complex chemistry means that while NO_2 levels can be controlled by limiting NO_x emissions, curbing ozone production may involve reductions in NO_x emissions, VOC emissions, or both. This is characterised by the notion of classifying a particular photochemical regime as 'NO_x sensitive' or 'VOC sensitive', which is described in detail by Sillman (1999), but the basic idea is best explained with reference to a particular type of diagram called an ozone isopleth. An example of such an isopleth diagram is shown in figure 1.1, which has NO_x and reactive organic gas (ROG) concentrations on the axes, and the resulting ozone concentrations appear as contour lines. For the purposes of the present discussion, the term ROG can be considered to be interchangeable with VOC. Looking at the plot it is clear to see that when NO_x is low, ozone concentrations are largely insensitive to changes in VOCs, and will increase with increasing NO_x. As NO_x increases further, the rate of ozone increase slows down, and for lower levels of VOCs, increasing NO_x reduces the ozone concentration. This gives a ridge-line on the plot which divides the two regimes, indicated by the black diagonal line. To the bottom and right of the line is the NO_x sensitive regime, where ozone levels are much more sensitive to changes in NO_x than VOCs. To the top and left of the line the opposite situation prevails; ozone levels are much more sensitive

to changes in VOCs, but are relatively unaffected by NO_x changes - this is the VOC sensitive regime.

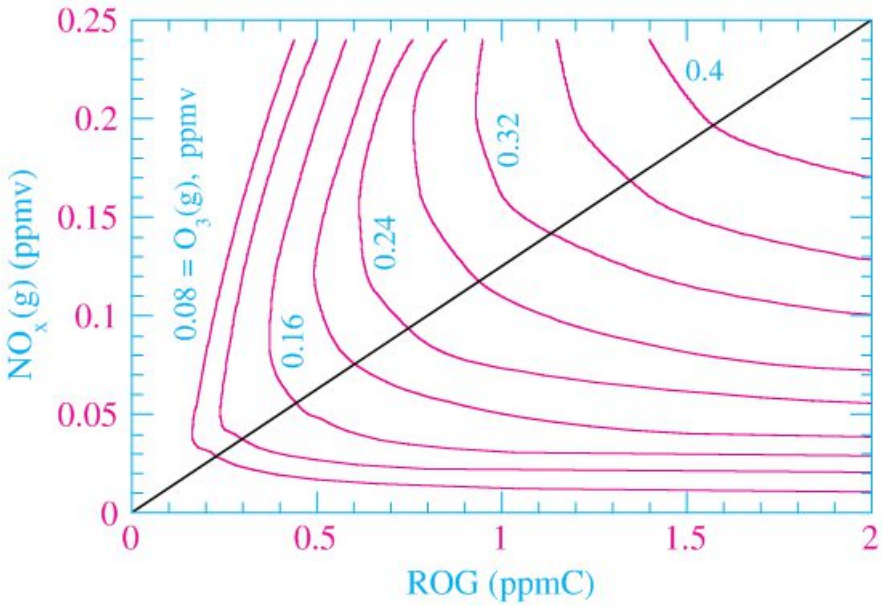


Figure 1.1: Ozone isopleth diagram showing resultant ozone concentrations for differing NO_x and ROG (VOC) mixtures. Taken from Jacobson (2002).

A severe UK ozone pollution episode in 2003 was described by Lee et al. (2006), which despite its extreme nature was typical of UK episodes in that it occurred during a period of very high temperatures (peaking at over 30°C) and anticyclonic conditions bringing air masses from mainland Europe. Lee et al. state that despite UK NO_x and VOC emissions being steadily reduced there has been only 'modest' impact on ozone levels. In a much more recent paper, Strong et al. (2013) study the same episode and agree that it 'occurred despite significant reductions in ozone precursor emissions across Europe since 1990'. Strong et al. conducted a modelling study of the episode and concluded that in order to reduce ozone levels during such an event different emissions control strategies would be required in different regions and even at different times during the episode.

The notion of ozone concentrations being sensitive to different emissions reduction strategies in different scenarios provides an opportune link to the next section, which concerns air quality legislation.

1.3 Air Quality Legislation

In response to the evidence on the adverse health effects of poor air quality, legislation has been introduced in many countries in an attempt to control the ambient concentrations of a variety of pollutants. In the UK the Clean Air Act of 1956 limited smoke emissions and was followed by the 1968 Clean Air Act which mandated the use of tall chimneys by industries burning fossil fuels. The UK joined the EU in 1972, and since that time Air quality legislation in the UK has been derived from EU directives. Directive 2008/50/EC 'on ambient air quality and cleaner air for Europe' (EU, 2008) established limits for the concentrations of a variety of pollutants, along with data quality standards to be used in their measurement. This directive was incorporated into UK legislation as 'The Air Quality Standards Regulations 2010', which gives limit values for NO_2 of 40 μg m^{-3} as a yearly average, and a 200 μg m^{-3} hourly average not to be exceeded more than 18 times per year. The regulations do not have a limit value for ozone, as it is a secondary pollutant, but instead have a target value for the maximum eight hour rolling mean of 120 μg m^{-3} not to be exceeded on more than 25 days per year averaged over three years. Exceeding the target value should be avoided by taking 'all necessary measures not entailing disproportionate costs' according to the regulations. This seems somewhat unsatisfactory, as one person's opinion as to what constitutes disproportionate costs is likely to differ from another's, especially if that person belongs to a high risk group with regards to possible adverse health effects.

In addition to limit values and target values for the atmospheric concentrations of pollutants, legislation also exists concerning maximum values for their emissions. The United Nations Economic Commission for Europe (UNECE) Gothenburg Protocol set limits for emissions of NO_x, sulphur dioxide, ammonia and VOCs which were to be met by 2010, and was amended in 2012 with new targets to be met by 2020. Having signed the original Gothenburg Protocol, the EU member states produced Directive 2001/81/EC 'on national emissions ceilings for

certain atmospheric pollutants' (EU, 2001), which in turn became UK law in the 'National Emission Ceilings Regulations 2002'.

As a result of the legislation outlined above, policy makers are required to produce strategies to control pollutant levels, and models can be used to assess the effectiveness of such strategies ahead of time, allowing cost - benefit analyses to set the cost of the abatement strategy against the potential improvement in air quality (Derwent et al., 2010). Another role for air quality models which arises when ozone target values are exceeded, can be to assess whether the current regime is NO_x or VOC sensitive, and so indicate the most efficient emissions reduction strategy.

Determining pollution levels requires a large amount of statutory monitoring at fixed sites. In a report written for DEFRA, Ferguson and Harrison (2010) state that in order to keep the number of monitoring stations required to a minimum, observational data is supplemented by numerical modelling to generate maps of annual mean pollutant concentrations, and other relevant information - this is another role played by models such as CMAQ. The same report points out that there are no formal criteria for quantifying the uncertainty in model outputs, and that the complexity of air quality models makes its estimation extremely difficult. In this context the quantification of modelling uncertainty would be valuable for policy makers, as it would enable the probability of the model output being correct to be included in the decision making process.

1.4 The Community Multiscale Air Quality Model

The development of CMAQ has been continuously funded by the United States Environmental Protection Agency (USEPA), with a philosophy that is demonstrated by the first two letters of its title; the C for 'community', reflects the fact that the model is open source and contributions to its development are made by

the wider scientific community. The M for 'multiscale' denotes that the model is designed to be used over a wide range of spatial scales. In line with the generally open nature of publicly funded science in the US, a great deal of information on, and source code for, CMAQ and related models can be downloaded from the Community Modelling and Analysis System (CMAS) website (CMAS, 2014). The model is designed to be run on a Linux computer platform and as the source code is compiled by the user there is considerable flexibility in the hardware which can be used. In the work described here it was compiled and run on a 108 processor cluster with a centOS operating system.

In common with other Eulerian CTMs, CMAQ solves a set of continuity equations, one for each modelled chemical species in each grid cell, at each time step of the simulated time period. The continuity equations have the general form (Jacob, 1999),

$$\frac{\partial n}{\partial t} = -\nabla \cdot (n\mathbf{U}) + P - L \tag{1.9}$$

where n is the number of molecules of the species in question and $\mathbf{U} = (u, v, w)$ is the wind velocity vector, so that $-\nabla \cdot (n\mathbf{U})$ is the divergence of the wind field times the number of molecules, and as such represents the difference in the number of molecules transported into and out of the grid cell. P is the sum of all sources of the species, for example emissions, and L is the sum of all sinks, such as deposition or photolysis. As transport out of one grid cell is equal to transport into the adjacent cell, these equations ensure that mass is conserved throughout the three dimensional model domain.

Many models involving atmospheric dynamics, particularly those operating over large spatial scales, make the assumption that the earth's atmosphere is non-compressible and is in hydrostatic balance, i.e. vertical acceleration of air in comparison to gravitational acceleration is small enough to be ignored. CMAQ is

designed to operate at spatial scales small enough to simulate urban air quality, and here the hydrostatic assumption must be abandoned if atmospheric dynamics are to be accurately simulated. In two papers Byun (1999a,b) describes the development of a set of governing equations for use in atmospheric dynamics for air quality applications, which address mass conservation problems caused by both non-hydrostatic dynamics and changing coordinate systems in multiscale scenarios. It is these equations which form the basis of the CMAQ model (Byun and Schere, 2006).

The general form of the CMAQ modelling system is shown in figure 1.2, where meteorological inputs are provided by the Weather Research and Forecasting (WRF) model, and the Sparse Matrix Operator Kernel Emissions (SMOKE) is responsible for processing emissions data into a form which can be input to the model. The Meteorology Chemistry Interface Processor (MCIP) ingests the meteorological fields produced by WRF and converts them into a form which is suitable for use by CMAQ and SMOKE, including performing any coordinate transformations which may be required (Otte and Pleim, 2010).

The initial conditions processor (ICON) can be compiled to run in one of two ways, depending on whether the model run is to be started from scratch or is a restart of a previous run. In the former case the initial conditions (ICs) are supplied to ICON by the user, as a text file containing concentrations of chemical species across the model domain, and in the latter case ICON reads output from a previous CMAQ run. A 'restart' can mean either running the next period of simulated time, or running a higher resolution domain which is 'nested' inside a coarser one. The boundary conditions processor (BCON) works in a similar way, either producing time invariant concentrations of all of the modelled species at the domain boundaries from a user supplied text file, or dynamic boundary conditions (BCs) from a CMAQ output file, which are used for nested domains. This will become clearer when the nested domains used in the model runs in this study are described shortly.

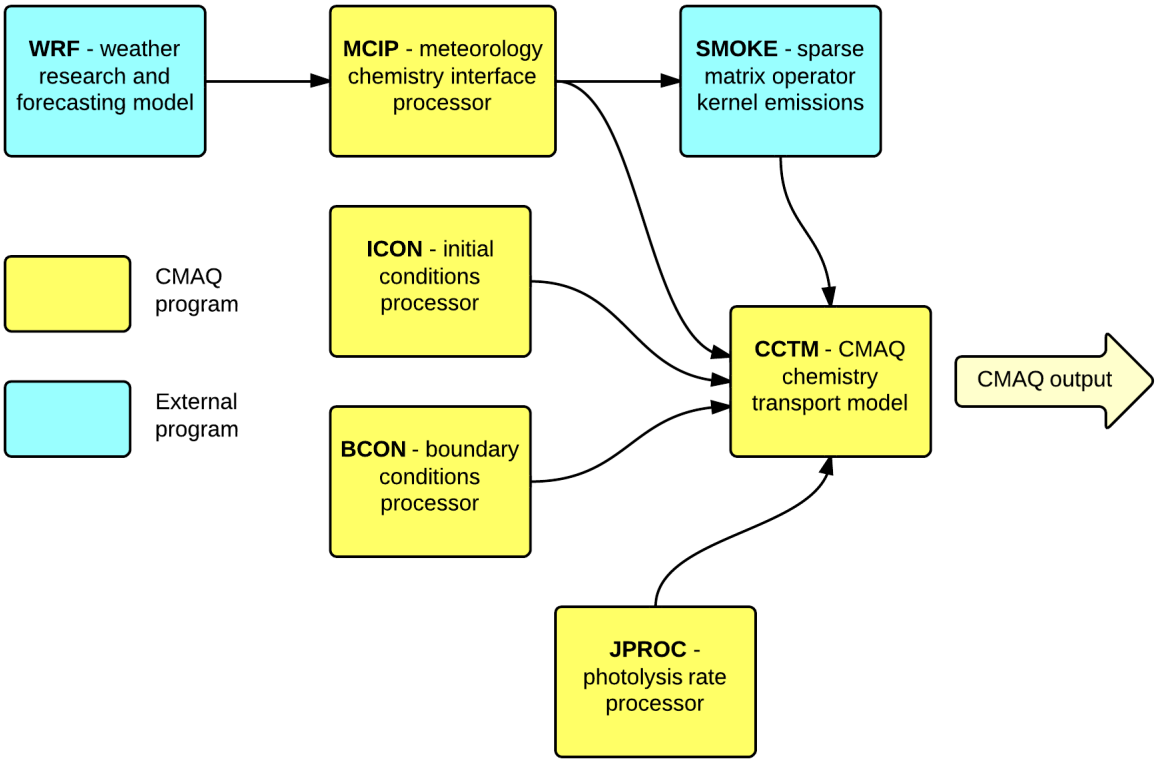


Figure 1.2: The flow of information though the various modules of the CMAQ modelling system

The Photolysis rate processor, JPROC, only needs to be run once for a given modelling scenario as it produces a set of look-up tables specific to the spatial domain and time period. These tables contain clear sky photolysis rates for all modelled species which undergo photodissociation, at each grid cell and time step. These photolysis rates are then modified during a particular run according to the amount of cloud cover over each model grid cell.

Finally, the CMAQ Chemistry-Transport Model (CCTM) does the major work of solving the model equations and producing the output files. The component of a CTM which contains the equations describing the modelled chemical reactions is generally termed the 'chemical mechanism', and the chemical mechanism used by CMAQ is known as Carbon Bond Five (CB05), as it was developed in 2005 (Yarwood et al., 2005). CB05 was, at the time of release of CMAQ v4.7.1, the latest in a series of mechanisms starting with the original carbon bond mechanism

described by Whitten et al. (1980). Chemical mechanisms designed for use in models such as CMAQ, which may need to be run on an operational basis as well as for pure research purposes, usually employ some kind of 'lumping' in order to reduce the number of chemical species and reactions which are represented and therefore reduce the computational expense of running the model. The carbon bond approach to lumping is to represent organic molecules by the type of bonds involving their carbon atoms. For example, the three carbon atoms in propene (C_3H_6) are joined by one single bond and one double bond, so 1 ppb of propene is represented in the mechanism by 1 ppb of single bonds and 1 ppb of double bonds (Whitten et al., 1980). This negates the need to represent propene explicitly in the chemical mechanism and instead it is represented by the two species 'PAR' for paraffin (single) bonds and 'OLE' for terminal olefin (double) carbon bonds. In the same way these surrogate species can also represent many other organic molecules. Exactly which species are lumped and which are represented explicitly varies from one version of the carbon bond mechanism to the next, but a full list of the species in CB05 can be found in Yarwood et al. (2005), which also lists all of the reactions and provides the formulae used to calculate reaction rates. A description of all of the input and output files, together with basic information on compiling and running the model can be found in the CMAQ operational guidance document (CMAS, 2010).

1.4.1 The Base-case model run

The model domain used in this project is shown in figure 1.3. It consists of an outer domain covering much of Europe, which has a grid cell size of 81 km. The innermost domain, which produces the model output of primary interest, covers the UK with a grid cell size of 9 km. Between these two is an intermediate domain with a grid cell size of 27 km, which aids in stepping down from 81 to 9 km resolution.

The modelled time periods are from 2006, chosen because a number of mod-

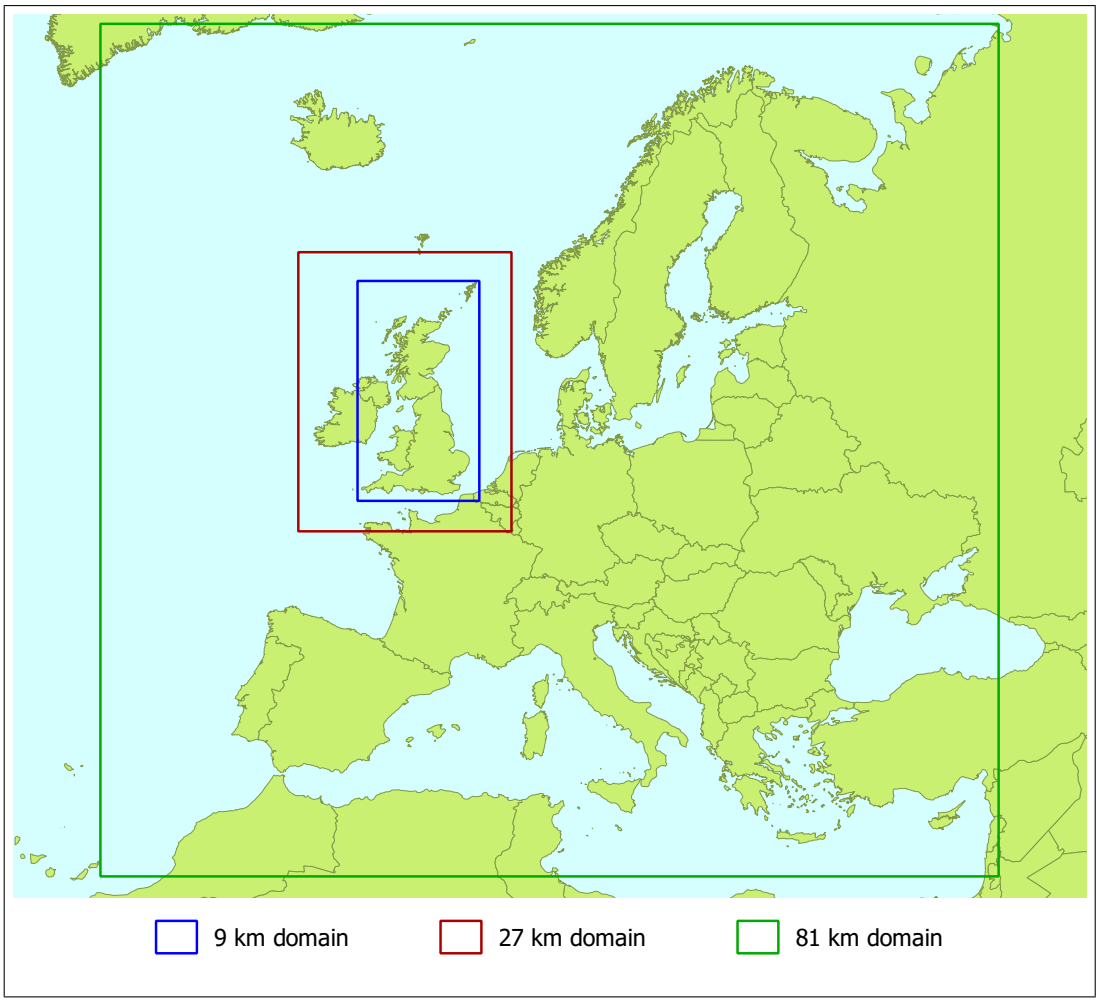


Figure 1.3: The outer, middle and inner model domains used by CMAQ in this project

elling studies and model assessment exercises have already been performed for this year (Carslaw et al., 2013; Rao et al., 2011), and so the present work should complement this. This also made 2006 a good choice for pragmatic reasons, as model-ready emissions and meteorological data were already available within the Environmental Research Group (ERG) at King's College London, and were kindly supplied by Dr. Nutthida Kitwiroon. The CMAQ model runs performed by Dr. Kitwiroon and described in the DEFRA report by Carslaw et al. (2013) covered the whole of 2006 and included several different emissions scenarios.The boundary conditions used for the model runs were derived from data provided as part of the Air Quality Model Evaluation International Initiative (AQMEII) (Rao et al., 2011). The emissions data for the outer (81 km) and middle (27 km)

domains was from the Euopean Monitoring and Evaluation Programme (EMEP) and those for the inner (9 km) domain from the National Atmospheric Emissions Inventory (NAEI). This emissions inventory is compiled on a yearly basis, by the combination of activity data from processes such as power generation, industrial activity and transport, with emissions factors defining the amount of a particular pollutant produced by each activity. The NAEI programme is responsible for submitting UK emissions data to the EU Nation Emissions Ceiling Directive (NECD) and the Convention on Long Range Transboundary Air Pollution (CLRTAP), and, 'in principle, ... makes estimates of all known emissions to air at as high a level of disaggregation as is possible' (Passant et al., 2014).

Focussing on the July 2006 pollution episode here allowed a much greater number of CMAQ runs to be carried out. An example of the model output during this period is given in figure 1.4, which shows ozone concentrations in the bottom layer of the model domain averaged over a one hour period. The vertical layers in the model are defined as constant pressure surfaces, so do not have a fixed height but vary in thickness with ambient air pressures. This bottom layer, which is the layer of most interest in terms of air quality, varies in height between about 10 and 15 m.

It is important to stress the computationally expensive nature of the model, and the large volume of input and output data involved in its execution. The time taken to perform a model run is highly variable and is dependent on the size of the domain, the length of the modelled time period and the hardware used, but it is illustrative to give a 'ball park' figure for reference in the discussions which follow. For the domains shown in figure 1.3 and a ten day modelled time period, it would take around 24 hours to run ICON, BCON and CCTM if only a single processor were used. It is common to to run the model on more than one processor, but the benefits of doing so scale quite poorly; two processors require considerably more than half the time of one, four processors considerably more then half the time of two, and so on.

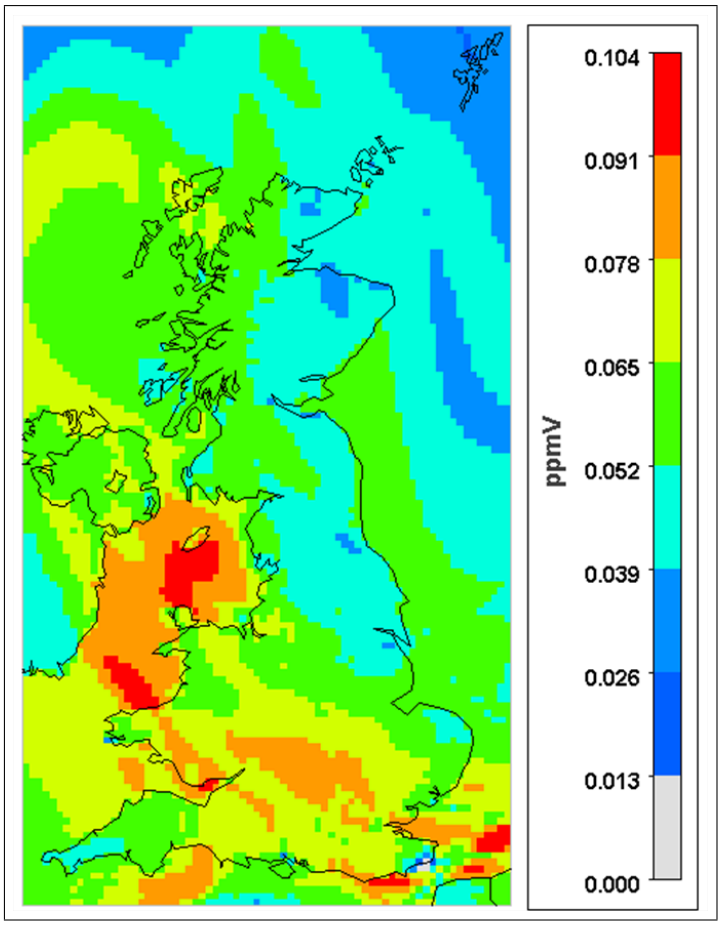


Figure 1.4: Example CMAQ output of peak afternoon ozone concentrations, 19th July 2006

1.5 Sensitivity and Uncertainty Analysis

At this point it is worth taking a moment to introduce some terminology which is common throughout the literature on uncertainty in deterministic computer models. Deterministic in this context means that a model comprises a set of equations which are formulated in a mechanistic manner with no stochastic elements, so that if it is run several times with the same set of input values it will produce identical output every time. Uncertainty in this output can arise from the fact that there is uncertainty concerning the values of the inputs, or from the fact that the model is an imperfect representation of reality. The input uncertainty can be referred to as input variable or input factor uncertainty, especially where those inputs vary over space or time, or parametric uncertainty, which often refers to model inputs which remain constant throughout a particular run.

This leads to the notion of the combined ranges over which these inputs may vary being defined as the input variable space, input factor space or parameter space. The terms will be used interchangeably here. The uncertainty which is due to a models' poor representation of some physical or chemical process, is usually referred to as model discrepancy or structural uncertainty.

Uncertainty analysis may just involve estimating the magnitude of uncertainty in model outputs, or may entail a more in-depth investigation of the relationships between the input and output uncertainties. This latter case is where uncertainty analysis converges with sensitivity analysis, which in its most basic form, involves changing the values of one or more model inputs or parameters, and assessing the impact on model outputs. This can be performed in several different contexts depending upon what is perceived to be the unknown quantity:

1. The true value of the model input is assumed to be known, but the effect of a change in its value on the real system is not.

2. The true response of the system to a change in the input is assumed to be known, and the model is being tested in order to assess whether it can reproduce this, sometimes called a dynamic evaluation.

3. There is uncertainty in the value of the model inputs, and the relative importance of those uncertainties is unknown.

The latter case is where the main focus of this report lies, however, the literature contains a great number of examples of all three cases, some of which are described in the next section. Sensitivity analysis can also be a much more rigorous process involving perturbations of all of the model inputs at once, performed in a more formal mathematical framework. Such an analysis has never been accomplished with the CMAQ model, and hence is one of the primary aims of this project.

Before moving on, a few definitions will be outlined which should help to frame the literature discussed in the following sections in the general context of

uncertainty analysis. All of the concepts here will be described in greater detail later on, so references are not given in this section, but rather are left until a particular subject is described in full.

The term Monte Carlo sampling is used in many areas and has become a general term for many kinds of random sampling. In the field of computer experiments, Monte Carlo uncertainty analysis involves assigning probability distributions to each of the uncertain model inputs, and then taking a random sample from their joint distribution, so that each point in the sample determines a particular value for each of the inputs. The model is run at each of these points, which propagates the input uncertainties through to the output, and gives a probability distribution around the output of the base-case run. In order to achieve a stable estimate of this distribution, a large sample size is required, and therefore a large number of model runs must be made. This process can be made more efficient by using a stratified sampling plan as an alternative to simple random sampling, such as Latin hypercube sampling (LHS), which will be expanded upon in due course, but for now can be thought of as choosing a given number of sample points in a way which will give more information than the same number chosen completely at random. If observational data are available which correspond to the model output they may be brought into a Monte Carlo analysis using one of a number of methods which weight the individual members of the sample according to the proximity of the model output to the observational data. Such methods include Generalised Likelihood Uncertainty Analysis (GLUE) and Bayesian Monte Carlo (BMC), both of which are simple to implement and allow the observational data to modify the shape of the distribution describing the model output. This is sometimes termed 'model calibration', and Bayes theorem, which allows the combination of data from different sources in a probabilistic framework, provides the natural vehicle for this kind of exercise. However, in practical terms, fully rigorous Bayesian analyses of complex environmental models with many uncertainties are often intractable, and such methods have probably seen more use in

computer modelling in engineering, where models are sometimes simpler, see for example Bayarri et al. (2007).

1.6 Sensitivity analysis in Air Quality Modelling

This section starts by looking at previous work which has been carried out with air quality models, focussing on CMAQ where possible, but citing work with similar models where appropriate.

1.6.1 Simple Sensitivity Tests

As part of a larger model evaluation exercise of CMAQ version 4.5 Appel et al. (2007) tested sensitivity of ozone predictions to synoptic meteorological conditions, boundary conditions, vertical resolution and choice of chemical mechanism. Their simulation covered the Eastern US at a 12 km grid resolution for the year 2001. Testing the model performance against observational data for the whole year they found that the model tended to under predict when observed ozone concentrations were high, and over predict when they were low, in other words it did not capture the temporal variability of ozone concentrations. The aim of the authors' sensitivity testing was to identify the causes of these errors. The approach taken to assess the importance of meteorology was to group the days from May to September into five synoptic clusters, based on mean sea level pressure patterns, and to calculate performance evaluation statistics for each one. Differences between the clusters were found, particularly in terms of mean bias, and the authors conclude that the model is able to predict ozone concentrations more accurately under some synoptic conditions than others. In order to test the effect of boundary conditions, Appel et al. (2007) replaced those used for their full year model run, which came from the GEOS-CHEM global CTM, with the default BC files supplied with the CMAQ source code, and reran the model for July of the same year. Performance deteriorated with the default BCs, which is

unsurprising as they are not specific to the modelled time period. The month of July was modelled again, this time using a higher vertical resolution, as the original model run did not use all of the vertical layers from the meteorological input data, but collapsed this from 34 down to 14 layers. When running with the higher vertical resolution, the authors found that the model was better able to predict those observations which were above 85 ppb and below 35 ppb. Version 4.5 of CMAQ used the Carbon Bond 4 (CB-IV) chemical mechanism, or the Statewide Air Pollution Research Centre 1999 (SAPRC99) mechanism, the choice of which is made by the user when compiling the cctm. Later versions of CMAQ use the updated Carbon Bond 5 (CB-05) mechanism and Appel et al. (2007) incorporated this into CMAQ 4.5 in order to test it against CB-IV. This must have taken considerable effort, so it is unclear why the authors chose these two mechanisms and not SAPRC99 as well, especially as again just the month of July was studied. Daily maximum ozone concentrations were 7-8% higher with CB-05 than CB-IV and this meant that the underprediction of high concentrations was lessened, but the overprediction at low concentrations was made worse. Finally the version of CMAQ 4.5 with CB-05 was compared to CMAQ 4.6, which also uses CB-05. The major difference between these two model versions highlighted by Appel et al. is a new planetary boundary layer scheme (ACM2) in CMAQ 4.6, and this gave a small improvement in model performance. Overall, this study represents an approach to sensitivity analysis which involves changing the configuration of the modelling system in order to find the best general configuration for ozone prediction in a particular domain.

In a similar vein, Appel et al. (2010) tested sensitivity to meteorological inputs by running two CMAQ simulations which were identical except that one used WRF to supply meteorological fields, and the other used the Fifth Generation Mesoscale Model (MM5). The approach taken was to compare the two simulations with observational data and to attempt to explain qualitatively any performance differences which were apparent. For example, higher summer ozone

concentrations in the WRF driven simulation are suggested to be likely to be due to smaller cloud fraction values (a number between 0 an 1 which represents the proportion of each grid square covered by cloud). However the authors are unable to determine whether the underlying mechanism was increased photolysis, or higher near surface air temperatures leading to greater biogenic VOC emissions, or a combination of both.

A more detailed study of the effects of changing the chemical mechanism from CB-IV to CB-05 was made by Sarwar et al. (2008), with a very similar CMAQ simulation to Appel et al. (2007), but this time including a wintertime period. This was followed up by a study in which CB-IV, CB-05 and SAPRC99 were compared side by side (Luecken et al., 2008). All of these chemical mechanisms employ some form of lumping in order to reduce the number of calculations required at each model time step and therefore reduce overall complexity and run times, but this is done in a very different way in the Carbon Bond mechanisms to SAPRC99. The Carbon Bond mechanisms employ a lumped structure approach, in which organic molecules are represented as functional groups, whereas SAPRC99 uses a lumped molecule approach, where some organic species are represented as a similar or generalised molecule (Faraji et al., 2008). In this respect Luecken et al. (2008) provide a useful comparison, in enabling the difference between an older and newer functionally lumped mechanism to be measured against the difference between molecular and functional lumping. They note that at that time all three mechanisms were in use in the US for regulatory purposes, and pose the question of whether changing the chemical mechanism in a model could change pollution control measures. Interstingly, the study finds that differences in ozone concentrations between CB-IV and CB-05 are often greater than differences between CB-05 and SAPRC99, suggesting that using a mechanism which incorporates the most up to date knowledge of chemical processes can be more important than how those processes are parameterised.

In another study from the US Tang et al. (2009) investigate the sensitivity of

the National Air Quality Forecast Guidance (NAQFG) to changes in BCs. They performed CMAQ runs with BCs derived from three global CTMs, the Model for OZone And Related Tracers (MOZART), the Real-time Air Quality Modelling System (RAQMS), and the Global Forecast System (GFS). In addition, two runs were made with BCs derived from ozonesonde measurements, and these five model runs were compared with the NAQFG standard fixed BCs used in the operational forecast. The model runs are evaluated against ground based-measurements and against ozonesonde data. Of the 15 ozonesonde locations used in the study, 11 are close enough to the model domain boundaries to be used in the creation of BCs. The authors use data from all 15 locations in the evaluation of the model runs, when perhaps it would have been more appropriate just to use the four remaining locations. The modelled scenario was a 16 day period starting on 21st July 2006, during which no pollution incursions were expected into the domain. For this reason the results are somewhat inconclusive, although Tang et al. (2009) state that the performance improvements gained in using BCs from global models are 'promising', which would appear to be a logical conclusion. One can only assume that the time period was chosen because of the availability of the ozonesonde data, but a useful extension to the study would be to run a longer simulation with global CTM BCs and the fixed BCs, hopefully covering some period of time during which polluted air might be expected to be transported into the domain.

The response of model output to changing emissions scenarios has been the focus of widespread interest. For example Mueller and Mallard (2011) investigate the contributions of background, natural and anthropogenic emissions to air quality in the US. They define 'background' as not being influenced by emissions from within their model domain, which includes the US and parts of Canada and Mexico. In this sense the natural and anthropogenic emissions are those which are normally defined as emissions in a model run, and the background emissions are those which influence the values in the BC input files. Modelled scenarios included 'all natural', which contained no anthropogenic emissions, and

'total emissions', which included all emissions sources. Also, several other scenarios included taking the total emissions case and removing lightning or wildfires. Amongst the interesting results from this study, Mueller and Mallard found that wildfire emissions were a major factor contributing to ozone levels greater than 100 ppb, and that ozone peaks associated with natural NO_x emissions tend to be high but brief. This clearly illustrates how the results of this kind of study can be highly specific to the geographical area being modelled, as one would expect total emissions due to wildfires to be very different in different parts of the world.

In a study of CMAQ response to emission control scenarios of 50% reductions in NO_x and VOCs, Arnold and Dennis (2006) examined the model response when using three different chemical mechanisms, CB-IV, SAPRC99 and the Regional Acid Deposition Model version 2 (RADM2). They found that the different chemical mechanisms responded differently to changing emissions, namely that SAPRC99 was more responsive to the changes, but also that the response varied according to the emissions characteristics at different locations. Arnold and Dennis conclude that sensitivity tests involving different chemical mechanisms should always include quantifying the response of the model to changing emissions.

All of the literature discussed so far involves modelling studies of US domains. In an example from a European domain, Gonçalves et al. (2012) use sensitivity tests to identify possible causes of errors in CMAQ output. Under predictions of nitrous acid (HONO) concentrations are identified as having a potentially major impact on modelled pollution levels beacause of the role of HONO as a source of hydroxyl radical (OH). The authors use five different scenarios to test the effects of including different HONO emissions sources and an alternative scheme for heterogenous chemical HONO production to the default in CMAQ 4.7.1. They find that including traffic emissions with an increased $HONO/NO_x$ ratio, and HONO production by NO_2 hydrolysis on ground surfaces, gave the best agreement with measured concentrations. Also in spain, Jiménez et al. (2007) examine the effects of ICs and BCs on a model domain in the north east of the country with

particularly complex and varied terrain. They find that the influence of BCs is more profound near the domain boundaries, as one would expect, and also that it is greater at background locations than urban ones. Jiménez et al. also emphasize the importance of accurate BCs for domains covering complex terrains, however their domain is quite small, at 272 × 272 km^2, and so BCs might be influential no matter what the characteristics of the terrain.

Sensitivity studies with CMAQ for UK domains have been very limited in terms of the number of studies undertaken. As part of a performance evaluation of CMAQ for a summer 2001 photochemical pollution episode, when peak ozone concentrations were under predicted, Yu et al. (2008) tested the model's sensitivity to doubling of anthropogenic NO_x or VOC emissions. Doubling of NO_x emissions led to a decrease in ozone concentrations for some of the modelled time period, but an increase over parts of the domain at other times, which the authors propose was due to transport of ozone from Europe. The doubling of VOC emissions led to increased ozone in urban areas, which Yu et al. attribute to the urban photochemical regime being VOC limited because of high NO_x concentrations. This is a reasonable assertion, but the results could also be influenced by the fact that the spatially resolved emissions data required by CMAQ was likely to have had most of the anthropogenic VOC emissions in urban areas; if emissions values are very low in other areas doubling them is unlikely to have much of an impact.

The effect of varying ozone deposition in a UK domain was tested by Emberson et al. (2013), who incorporated the Deposition of Ozone for Stomatal Exchange (DO_3SE) model into CMAQ, allowing not only the model's sensitivity to deposition of ozone to vegetation to be examined, but also to estimate the ecosystem impact this causes. The authors particularly wanted to assess the effects of moisture conserving strategies by plants during hot dry weather, which reduce stomatal conductance, and consequently the uptake of ozone. In addition to the base case, which was run for the whole of 2006, two additional scenarios were modelled for June and July, a 'no stress' scenario in which stomatal con-

ductance was not limited by soil moisture deficit, and a 'stress' scenario in which stomatal conductance was fully limited. The authors compared the soil moisture deficit predicted by CMAQ with Met. Office data, as this was used by DO_3SE in its calculations of ozone deposition, and found good agreement. For June and July the number of days when 8 hour mean ozone concentration exceeded 100 μg m^{-3} was 20 for the base case, 24 for the stress case and only 8 for the no stress case, showing a high degree of dependence on dry deposition when pollution levels are at their highest.

The literature discussed so far represents a small sample of the kind of work which has been carried out around the world which in some way or other attempts to address sensitivity issues within the CMAQ model. However, the kinds of sensitivity tests described, which involve a very limited number of scenarios, fall a long way short of what a statistician would describe as a full sensitivity analysis. This is partly a result of the long model run times associated with CMAQ and the amount of work often involved in setting up and running a simulation, and partly due to the modellers' desire to select values of input variables which have some kind of physical relevance.

1.6.2 Direct Sensitivity Analysis Methods

An attempt to overcome the aforementioned difficulties in performing sensitivity analyses with CMAQ has been the implementation of methods to compute local sensitivity coefficients directly, as the model is running. The term 'local' implies that the sensitivity of the output to a particular input is calculated over a small range of variation of that input, close to its baseline value and so is comparable to a partial derivative (Saltelli et al., 2008). These local sensitivity coefficients could of course be calculated in a brute force fashion, with just two model runs, each having a slightly different value of the input parameter in question, but this may become impracticable if many coefficients are desired. A solution is to implement the direct method with an auxiliary set of equations, which are derived from the

model equations and in some cases these are coupled and solved together (Dunker, 1984). Dunker goes on to describe how this process can be unstable, better results being obtained by decoupling the two sets of equations and solving them separately, for this he introduces the term Decoupled Direct Method (DDM). This technique was implemented in a three dimensional air quality model, the California/Carnegie Institute of Technology Airshed Model, by Yang et al. (1997) who refer to it as DDM-3D. Extending this idea to take into account non-linearity in the model's response to changes in input factors, the High-order Decoupled Direct Method (HDDM) was implemented by Hakami et al. (2003) to compute higher-order derivatives in the Multiscale Air Quality SImulation Platform (MAQSIP), a prototype of CMAQ (Mathur, 2005). Hakami et al. compare the HDDM first and second-order derivatives, which they refer to as first and second-order sensitivity coefficients, to brute force calculations of sensitivity when NO_x and VOC emissions are perturbed by $\pm$ 10 % , and find good agreement. The method can, in theory, calculate what Hakami et al. term 'cross sensitivities', that is sensitivity of the model output to two or more inputs being perturbed together, but the paper does not make it clear whether or not this was actually implemented. In applying this method to CMAQ, Cohan et al. (2005) call it the High-order Decoupled Direct Method in Three Dimensions (HDDM-3D), but in order to stem the proliferation of acronyms, all such methods will be referred to here as the DDM. Using CMAQ 4.3 with the SAPRC99 chemical mechanism, Cohan et al. claim to find good agreement between brute force calculations of first and second-order sensitivities, and those calculated with the DDM, for anthropogenic NO_x and VOC emissions. They calculate the sensitivities of 8 hour mean ozone concentrations for every grid cell and report domain wide mean values, which do indeed appear to be in good agreement. However, the normalised mean bias and error of the grid cells is also reported, and the normalised mean error in particular shows a different picture, with values ranging from 2.5 % to 49.2 %. These tests of the first and second order coefficients are made against brute force calculations

with emissions of ± 10 % of the base case. Cohan et al. (2005) go on to test the DDM coefficients against brute force calculations with 50 % and 100 % reductions in NO_x and VOC emissions by extrapolating the coefficients. The linear extrapolation of the first-order coefficients provides poor agreement with ozone levels calculated using the brute force method. When incorporating the second-order coefficients by using a second-order Taylor series expansion to predict ozone levels, as first suggested by Hakami et al. (2003), the agreement improves, but still produces errors of over 10 % in all cases but one. The authors use the 100 % reduction cases to compare the contribution of different emissions to ozone concentrations and include cross sensitivities in this process to provide an estimate of the interaction effect between NO_x and VOC emissions, to demonstrate how the method could be used in a source apportionment exercise. Also, they discuss how uncertainty in the emission inventory value of a particular species can affect the sensitivity coefficients, and provide a way of adjusting the coefficients for known errors, if for example a retrospective alteration was made to an emissions inventory. The method does not, however, have any way of accounting for uncertainties in all of the input factors at the same time. Interestingly, the paper has no conclusion or final discussion section, so the authors do not express their opinions on the accuracy of using extrapolated sensitivity coefficients, but in Cohan et al. (2006) the method is used to assess the cost effectiveness of various NO_x and VOC emission reduction strategies by weighing their effect on ozone concentration reductions against the monetary cost of implementation. Here, Cohan et al. (2006) cite Cohan et al. (2005), making the assertion that the Taylor series expansions 'accurately capture the concentration - emission response of the underlying model even for perturbations of 50 % or more'.

The ability to calculate the sensitivity of PM concentrations to emissions was added to the CMAQ DDM by Napelenok et al. (2006), followed by calculation of sensitivities to reaction rate constants (Napelenok et al., 2008). In evaluating these improvements against brute force calculations, Napelenok et al. (2008)

report better performance with respect to primary pollutants, than secondary pollutants, particularly in the case of secondary organic aerosol (SOA). They highlight the sensitivity of sulfate aerosol to ammonia emissions as being particularly poorly represented by the DDM, and point to the non-linearity of secondary relationships as the possible cause, as second-order sensitivity capability for PM had not been added at this stage.

The DDM was used by Jin et al. (2008) to assess the possible impact of different emissions reductions on ozone levels in the San Joaquin Valley, California. They use just the first-order sensitivity of ozone concentrations to NO_x and VOC emissions to determine which of these is most influential in different parts of their study domain, stating that first-order sensitivities are 'likely to be valid for emissions perturbations $\leq$ 25 %, which are usually suitable for policy applications'. Jin et al. provide no validation evidence in support of this assertion, but instead cite Vuilleumier et al. (1997). However, Vuilleumier et al. do state that first-order sensitivities are generally valid when parameter variations are limited to at most 25 %, but their study uses a single cell model which simulates atmospheric reaction chamber experiments, and as such does not even have emissions as input factors.

1.7 Uncertainty Analysis in Air Quality Modelling

1.7.1 Monte Carlo Analysis of CMAQ by Direct Methods

The Taylor series expansion of DDM coefficients, used to predict ozone concentrations at untried values of the input factors, was used as a surrogate for CMAQ in a Monte Carlo uncertainty analysis by Tian et al. (2010), who use the term Reduced Form Model (RFM) to describe the expansion. Lognormal uncertainty distributions were assigned to anthropogenic and biogenic VOCs, and point and area NO_x emissions sources, in such a way that 95 % of the probability mass fell within a given factor of the baseline value for that emission. Those factors ranged

from 1.5 to 3, and 10,000 random samples were taken from the joint distribution of those inputs and propagated through to ozone output concentrations using the RFM. This means that a large number of predicted ozone values were made with a RFM which is extrapolating the sensitivity coefficients to two, three or more times the baseline values. The authors justify this by citing Hakami et al. (2003) and Cohan et al. (2005), who do not actually make such claims for the accuracy of the DDM, and in any case, just because a RFM has been validated in one scenario, it does not automatically follow that the validation can be transferred to another scenario, especially when the method used to produce the RFM has extrapolation at its core. Tian et al. acknowledge this problem and try to circumvent it by characterising the induced output uncertainty distribution using its 'inferred coefficient of variance' (ICOV), which they define as the ratio of the inferred standard deviation (ISTD) and the median. They state that the (ISTD) is calculated using the $\pm$ 1 standard deviation confidence interval, and so is robust to outliers, but do not explain how this occurs. Fundamental to this method, is also the idea that removing outliers from the output distribution will automatically remove them from the input distributions, but it is not clear that this is the case, especially as the Taylor expansions used in the RFMs include cross sensitivities.

Monte Carlo uncertainty propagation with RFMs is taken a step further by Pinder et al. (2009), who use the technique in combination with an ensemble of CMAQ model structures. In common with many models which are compiled from source code by the end user, CMAQ has a number of science options which are chosen at compile time, and may affect the output of the model. CMAQ 4.5 was used with both the CB-IV and SAPRC99 chemical mechanisms, and in each case was compiled with three different combinations of land surface model and planetary boundary layer scheme, to make a total of six ensemble members. Pinder et al. create RFMs from the first and second-order sensitivity coefficients for ozone BCs and NO_x and VOC emissions, and the cross sensitivity for NO_x

and VOC emissions, of each of the six ensemble members. The Monte Carlo samples were made by randomly selecting one of the ensemble members and then randomly sampling emissions and BC values from uniform distributions, for VOCs for example from one half to one and a half times the base case value. The authors of this study also cite (Cohan et al., 2005) as justification for the RFM method, stating that it would reproduce CMAQ ozone values to within a few percent. In actual fact, Cohan et al. found that a 50 % VOC reduction gave an error of 10.5 % when comparing their RFMs to CMAQ output. In terms of emissions increases above the base case level, the maximum tested for any species by Cohan et al. was 10 %, and their study was just with SAPRC99 and for a different year. The output probability distribution does appear to perform quite well, however, when compared with observational data, as 42 % of observations fall within the interquartile range, and 96 % of observations fall between the highest and lowest ensemble members. Given that the authors have not included a comprehensive list of uncertain model inputs, this could indicate that the ranges they chose for those inputs were too large and thus compensated for other unaccounted for uncertainties.

If this section of the literature review (and indeed the previous section) appears to be very US - centric, it is because all of the published work using the DDM concerns US domains. Although DDM based methods have been strongly criticized here, they do have a place in sensitivity and uncertainty analysis, especially for exploratory work, as there is no other method which can provide similar information at such low computational expense. The criticism mainly centres on the use of RFMs which have not been validated to make important recommendations, such as those affecting public policy.

1.7.2 Monte Carlo Analysis of Other Air Quality Models

Here we move on to other air quality models, as the work with RFMs described above represents the only kind of uncertainty analyses which have been carried

out with CMAQ.

Ozone concentrations produced by the Urban Airshed Model version five (UAM-V) in a US domain were the subject of such a study by Hanna et al. (2001), in which they identified 128 uncertain input variables and sought expert opinions in order to estimate their uncertainty. A hundred experts were contacted and asked to fill in information on a web page, but ony about 20 responded, and the authors found that they obtained more information by meeting with groups of experts. Hanna et al. also point out that most of the experts thought that correlations must exist between some of the input variables, but these general opinions could not be transformed into a specific correlation structure. Therefore the study was carried out with the assumption of independence between variables. The list of variables included ICs and BCs, emissions, meteorological inputs, reaction rate coefficients and photolysis rates, most of which are assigned lognormal distributions, the authors stating that this is usual for most geophysical variables. Whilst this assertion is often true for the distribution of the absolute values of many environmental variables, it is not so clear that it should also be true for the uncertainty distributions of those values. Also, anthropogenic emissions are not really environmental variables, and common sense would dictate that the base case value of an emission is the best estimate of that emission, and so any uncertainty assigned to it should at least be symmetrical. It is conceivable that emissions could have skewed distributions, perhaps if a bias was thought to exist due to under-reporting of traffic emissions for political reasons, for example, but unless such a belief is stated explicitly it is more appropriate to assume a symmetrical distribution. 100 model runs were made with different randomly sampled sets of input variables for each of four scenarios; 1995 emissions, projected 2007 emissions, a 50% reduction in 2007 anthropogenic VOC emissions, and a 50% reduction in 2007 anthropogenic NO_x emissions. In addition to producing estimates of uncertainty in the absolute predicted values of ozone, Hanna et al. also produced regression and correlation based sensitivity measures, but state that

their confidence in these is low because of the small number of Monte Carlo runs made. However they do have confidence that 100 runs is adequate to correctly characterise the variance of the model output. This would seem to be doubtful, as it is well known that small samples taken from high dimensional input spaces (in this case 100 points in a 128 dimensional space) are unlikely to cover the space effectively enough to allow such inferences to be made (Santner et al., 2003). Hanna et al. (2001) point out that the Monte Carlo method they use only takes into account uncertainties in the model input factors, and does not account for any deficiencies in the model's representation of the real system, and also that those input uncertainty distributions can only be regarded as estimates. A final point made by Hanna et al. (2001), which is of great relevance to all sensitivity and uncertainty studies in air quality modelling, is that there is no satisfactory way to perturb meteorological fields without disturbing the mass balance upon which predictions of pollutant concentrations depend. In other words, if the wind speeds and directions, in particular, are perturbed then the difference between the mass entering and leaving a grid cell in a particular time step may not equal the change in mass in that grid cell, hence violating the continuity of equation 1.9.

In a Monte Carlo uncertainty analysis of a Lagrangian model, the California/ Carnegie Institute of Technology air quality model, Bergin et al. (1999) attempted to make more effective use of the number of model runs they performed by selecting the values of the input parameters using Latin Hypercube Sampling (LHS) (McKay et al., 1979). LHS is effectively a stratified sampling scheme which ensures more complete coverage of the input variable space than simple random sampling for any given number of sample points, and will be described in more detail in chapter 2. Lagrangian models differ from grid based models such as CMAQ, in that they simulate the chemistry in a single parcel of air as it moves across a domain, rather than simulating the chemistry throughout the entire domain. As well as propagating the input uncertainties through to

the model output, Bergin et al. (1999) also examine the sensitivity to each of the inputs using linear regression based analysis. As they point out, this can be unreliable in cases where the model output is non linear with respect to the input being perturbed, and scatter plots were produced to check the analysis. Bergin et al. do not show the plots, but say that they were produced for 'about twenty' variables, which is under half of the number of variables used in the study.

In an extension to the above study, Bergin and Milford (2000) enhance the simple Monte Carlo Analysis by using a Bayesian Monte Carlo (BMC) technique introduced by Dilks et al. (1992) in the field of water quality modelling. Bergin and Milford identify a problem with the previous analysis in that certain combinations of input factors produced by random sampling gave rise to model output which was a poor match to observational data. The solution suggested by Dilks et al. is to weight each of the Monte Carlo runs with its probability of being correct according to the error in the output when compared to observational data. They achieved this by noting that when the error in the observational data can be assumed to be normally distributed, a normal likelihood function can be used in a simple application of Bayes' theorem to weight the discrete set of model runs. Mathematical details are provided in chapter 2, but the salient point is that by weighting the individual model runs from the Monte Carlo analysis, both the input and output distributions can be calibrated in the same process (Dilks et al., 1992). As they were using a Lagrangian model, Bergin and Milford (2000) used the interpolated values of measured ozone concentrations at two locations, and used the error in the interpolation technique calculated in a previous study (McNair et al., 1996) as an estimate of the observational error. Also, they used five hour sections of each trajectory, along which the errors in the interpolated observations were serially correlated. Bergin and Milford modelled this correlation as a Markov process which enabled them to use a likelihood function derived from the bivariate normal probability density function. As a result of the calibration process, the means of the Monte Carlo output distributions were shifted

closer to the observed values, and their standard deviations were reduced. Also, six of the input parameter distributions had mean values changed by between five and fifteen percent, with the rest changed by less than three percent. Bergin and Milford point out that the BMC method assumes that all significant sources of uncertainty have been accounted for in the initial Monte Carlo analysis, and that a potential major source of error in the Lagrangian model they use is that it has no treatment of horizontal diffusion. Hence, they suggest that BMC should be applied with a Eulerian CTM. This suggestion is an interesting one in that it illustrates the way a modelling practitioner might be predisposed to address the issue of model discrepancy, that is to reduce it by using a more sophisticated air quality model, rather than to incorporate it into the statistical model used in the analysis.

This was done by Beekmann and Derognat (2003) with the CHIMERE model over the Ile-de-France region (the region around Paris) nested within a larger European domain. The analysis was performed for three days during July and August 1999, when airborne monitoring data from circular flights around the region was available in addition to the usual ground-based monitoring stations. These circular flight paths allowed Beekmann and Derognat to plot NO_y (NO_x + PAN + HNO_3 + HONO) as a function of direction from the centre of Paris and thereby identify the position of the urban plume. It was the pollution levels within the plume compared to background levels that the authors were interested in, so they subtracted the average of the lowest 30 % of measurements from the average of the highest 30 % of measurements for each complete circle of the flight path. This was done for the airborne measurements of NO_x, NO_y, and O_x (O_3+NO_2), and ground-based O_x measurements were also used in the BMC process. 500 Monte Carlo simulations were made by perturbing 26 model input parameters, and the in-plume minus out-of-plume subtraction was done for each model run so that the data would be equivalent. This procedure was an ingenious way of removing some unaccounted for sources of uncertainty from the analysis, such as

small scale variations in wind direction. Beekmann and Derognat found that the BMC greatly improved the model output agreement with observations, and of the modifications to input distributions, among the most notable was the mean of the NO_x emissions being shifted upwards by 30 % on one day and downwards by 21 % on the next.

Deguillaume et al. (2007) performed a very similar BMC experiment using the same model domain, but this time for the two summers of 1998 and 1999 using observational data from ground-based monitoring stations, and with the express aim of calibrating emissions. Measurements used were of NO and ozone from urban sites within Paris and from six rural sites in the surrounding area, which allowed characterisation of the urban plume in a similar way to the previous study. This time the BMC prodedure was carried out with long-term average measurements and with an initial Monte Carlo sample of 500 model runs. In this case the mean of the NO_x emissions was not shifted greatly, but the distribution was sharpened, with the variance decreased by around 60 %. The authors conclude that the shorter study (Beekmann and Derognat, 2003) allowed no inference about bias in NO_x emissions to be made, but that this longer study demonstrates an absence of bias.

The same data was augmented with extra model runs in a following study (Deguillaume et al., 2008) in order to investigate whether ozone production in the Paris area is NO_x sensitive or VOC sensitive. From the 500 Monte Carlo model runs performed previously, the 100 runs which showed the best match to observations after the BMC were selected for use in emissions reduction scenarios, in which either NO_x or anthropogenic VOCs were reduced by 30 %. For reduced NO_x emissions compared to the base case, the authors found increases in ozone on some days and very little change on others, with the increases more pronounced within the urban area due to reduced ozone titration. The decreased VOC emission scenario tended to give reductions in peak ozone levels, particularly in the urban plume and slightly less so in the urban area itself. Comparing

the two scenarios to each other, by subtracting ozone levels in the reduced VOC scenario from the reduced NO_x scenario, Deguillaume et al. find a VOC sensitive regime 60 % of the time, 'neutral behaviour' 40 % of the time and they report that the area was never NO_x sensitive during the study period. The authors go on to repeat the BMC approach with the emissions reduction scenarios in order to test their findings in the light of model input uncertainty, and conclude that the 10th and 90th percentiles of the ozone output distributions show largely the same picture, except that the geographical extent of the VOC limited regime is uncertain. In their concluding remarks Deguillaume et al. state, in agreement with other authors e.g. Dilks et al. (1992), that the BMC method only takes into account uncertainty in the model inputs but not in model parameterisations. They go on to say, however, that this parameterisation uncertainty is implicit in the input uncertainties. It is not clear exactly what is meant by this, but if the authors mean that model structural uncertainty can be subsumed into the input uncertainty, this would seem to be a bold statement for a complex Eulerian CTM.

A model which is similar to CMAQ, the Comprehensive Air Quality Model with extensions (CAMx) (ENVIRON, 2010), was used by Digar et al. (2013) in a study which compared model calibration using BMC and a further two methods. The initial MonteCarlo procedure was similar to that performed by Pinder et al. (2009) in the uncertainty analysis of CMAQ described earlier, but was for a domain covering part of Texas for 2006. The DDM has been implemented in CAMx which allowed Digar et al. to produce Taylor series based RFMs of ozone response to changes in various model inputs, including emissions of NO_x and anthropogenic and biogenic VOCs, plus several reaction rates and model domain BCs of a number of species. This was done for an ensemble of four model structures, which included the standard model configuration, an alternative chemical mechanism, alternative biogenic emissions and a combination of these latter two. 1000 samples were made with the RFMs for each ensemble member, giving a total of 4000 Monte Carlo samples. In common with some of the work reviewed

earlier using RFMs, they are not validated against brute force model runs, which devalues what is otherwise quite a comprehensive uncertainty analysis. Model outputs were weighted using BMC with ozone and NO_x measurements from a ground-based monitoring network during a month long pollution episode. The measurements were averaged over time at each site in an effort to compensate for day to day meteorological errors but still allow spatial comparisons. In a second calibration method the ozone measurements were used to evaluate each Monte Carlo run against the standards set by the USEPA for air quality models to be used for regulatory purposes. In a simple binary process model runs were deemed either acceptable or unacceptable, and Digar et al. assigned equal weights to all of the acceptable runs to produce adjusted distributions. A third method also involved a binary screening process, but this time the acceptability criterion was based on a nonparametric test, the Cramér-von Mises two sample test (Anderson, 1962) of whether two samples are drawn from the same underlying distribution. This was used with the distribution of measurements and the distribution of model outputs for each Monte Carlo run, to again give an adjusted but equally weighted distribution of Monte Carlo runs. When comparing the three methods Digar et al. find that the means of the distributions of model output obtained using the two screening based methods show a better fit to observational data than than did the mean of the BMC distribution. They suggest that this was because more of the observational data was used in the two screening methods than in the BMC. This hardly seems like a surprising conclusion given that the authors used the same data for the calibration processes and its evaluation.

The BMC method and the other calibration methods used by Digar et al. share much in common with Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley, 1992), in that all of these methods involve taking an initial Monte Carlo sample in which all of the model runs are considered equally likely, and then post processing that sample according to some kind of comparison of the model output with observational data. The 'generalized likelihood' in GLUE

is described by Beven and Binley as 'a fuzzy, belief, or possibilistic measure of how well the model conforms to the observed behaviour of the system'. If one of the binary screening methods described by Digar et al. had been combined with the BMC to weight the accepted model runs, then this would indeed be a particular application of the GLUE methodology. In GLUE the most unlikely runs are ruled out and those remaining have their likelihoods rescaled to sum to one (Beven and Binley, 1992), as would be the case had they been generated by a formal likelihood function. GLUE has proved to be an extremely popular method within the hydrological modelling community, but has also attracted criticism, as the probability bounds assigned to model predictions with a generalized likelihood are often very different to those which would be realized using a formal likelihood function (Stedinger et al., 2008). This criticism could perhaps be seen as not precluding the use of the method entirely, but as a reminder that whenever non-standard methods are employed the results must be interpreted with care. Stedinger et al. conclude with the remark that GLUE will produce results which are consistent with both classical and Bayesian statistics if an appropriate formal likelihood measure is used.

Derwent and Murrells (2013) use a method which they describe as a simplified version of GLUE in a study which is designed to estimate the contribution of chemical mechanism choice to uncertainty in modelled response to emissions reductions. To achieve this the authors use a Lagrangian model into which they can substitute various chemical mechanisms, and evaluate the impact of this on VOC and NO_x emissions reductions, for trajectories which arrive in the UK during July 2006. The standard formulation of the model used the Common Representative Intermediates mechanism Version 2 (CRIv2) which is a more complex mechanism than those which have been mentioned earlier in this review, comprising 1168 chemical reactions (Derwent and Murrells, 2013), whereas CB05, for example, has 156 reactions. The authors performed a Monte Carlo analysis with this version of the model in which various emissions, BCs and reaction rates were per-

turbed. Also among the list of variables were meteorological factors; temperature, relative humidity, boundary layer height and the latitude and longitude of the air parcel. As the Lagrangian model only has one grid cell, being effectively a moving box, meteorological variables can be perturbed without causing the mass balance issues that would be a problem with Eulerian models. Derwent and Murrells performed 1000 model runs for each day of July 2006, making a total of 31,000 runs, and the predicted ozone concentrations were compared with measurements at Harwell in Oxfordshire. 9006 runs were found to have output within ± 5 % of the measured data and these were deemed to be acceptable. The remaining runs were discarded from the analysis. The 9006 acceptable runs were repeated with a 30 % reduction in NO_x emissions and a 30 % reduction in VOC emissions, and these were categorised as NO_x sensitive or VOC sensitive depending on whether the NO_x reduction or the the VOC reduction resulted in lower ozone predictions. By plotting the change in ozone concentration caused by emissions changes against the ratio of ozone to NO_z in the case of no emission reduction, where $NO_z = HNO_3 + PAN + \text{particulate nitrate} + 2\,N_2O_5$, Sillman (1999) demonstrated that a threshold exists in the ozone/NO_z ratio below which the regime is VOC sensitive and above which it is NO_x sensitive. By demonstrating that this behaviour occurred in their 'acceptable' sample of model runs Derwent and Murrells concluded that the Lagrangian model with the CRIv2 mechanism had passed a diagnostic evaluation for response to emission changes. This test does not however place any emphasis on the magnitude of the model response. The whole process was repeated with four different chemical mechanisms, CB-IV, CB-05, SAPRC-99 and an updated version, SAPRC-07. In these cases the authors define as acceptable all those model runs which were within ± 3σ of observations, but do not say how this quantity relates to the ± 5 % used earlier. The 50th percentile of the ozone response to NO_x reductions was compared to the 50th percentile of ozone response to VOC reductions for the accepted Monte Carlo runs, in order to classify each day as NO_x or VOC sensitive for each chemical mechanism. Derwent and

Murrells report that all of the mechanisms were in agreement on 24 of the days, but there was some disagreement on the remaining 7 days. Examining these 7 days further, however, they state that the differences in the 50th percentile values are small compared to the spread of the Monte Carlo output, and illustrate this with box plots for one of the days, concluding that the uncertainty in ozone predictions caused by choice of chemical mechanism is insignificant when other causes of uncertainty are taken into account. Unfortunately the authors chose to anonymise the mechanisms in their results, not revealing whether the differences were greater between different lumping methods or between older and more up to date mechanisms.

The above methodology is placed on a more sound theoretical footing in the process of 'history matching' (Craig et al., 1996), which is similar in that it attempts to reduce the volume of the input variable space by taking an initial Monte Carlo sample and ruling individual members in or out according to a comparison with observational data. Craig et al. used a simplified Bayesian approach, Bayes linear, where probability distributions are specified only up to their second moments (with means, variances and covariances), to formalise the ad-hoc calibration of models used in the oil and gas extraction industry. The inefficiency of these types of methods in general, in terms of the small proportion of model runs which may be found to be acceptable (Stedinger et al., 2008) has meant that model surrogates have often been used in place of the deterministic model. For example, history matching with a quick to run surrogate for a computationally expensive climate model was employed by Williamson et al. (2013), who give the sparsity with which the input space can be sampled using the original expensive model as another reason for using a surrogate. They state that using a few runs of the expensive model, and ruling out those which are not a good match to observations, also means ruling out unsampled areas of the variable space around the rejected points. At this point it is worth taking a brief aside, before returning to history matching, to consider the idea of model surrogates, as

much of the remaining literature relies heavily on their use. The RFMs produced using Taylor series expansions of DDM coefficients discussed earlier would also fall into this category, but have been considered separately here, as they differ from other methods in terms of the considerable extrapolations involved in their construction.

1.7.3 Metamodelling of Computationally Expensive Models

Here, the term 'metamodel' will be used instead of surrogate, as sometimes producing the metamodel could be an aim in itself, if for example it was simple enough to allow direct inferences about the real model to be made (Fang et al., 2006). Fang et al. describe a number of methods for producing metamodels which are common in statistical modelling, including polynomial regression, splines, kriging and neural networks. All of these methods require the computer model to be run a number of times to provide training data, in contrast to the RFMs, which only use a single model run.

The Stochastic Response Surface Method (SRSM), developed by Isukapalli et al. (1998), is a metamodelling technique which employs polynomial chaos expansions to express a models' output uncertainty as a function of it's uncertain inputs. Polynomial chaos is a general procedure for representing a random variable as a function of another random variable or variables and so provides a suitable vehicle for describing uncertainty in the model output when the inputs are expressed as random variables. The model is run at various values of the input variables in order to provide training data to determine the values of the polynomial coefficients. Isukapalli et al. state that the number of model runs required to deduce the appropriate values of the polynomial coefficients is significantly less than the number of model runs that would be required in, for example, a Monte Carlo uncertainty analysis of the original model. In the same paper they apply the SRSM to a Lagrangian model of the dispersal of industrial atmospheric

emissions, the Reactive Plume Model, version 4 (RPM-IV), which uses the CB-IV chemical mechanism. In evaluating the SRSM they carry out a 10,000 sample Monte-carlo analysis of the model with which to compare their response surface, and require 600 training runs in order to achieve acceptable results with nine uncertain input parameters. The reason for describing this technique in particular, from amongst the plethora of possible metamodelling techniques, is that it has been proposed for use in the uncertainty analysis of CMAQ (Isukapalli et al., 2005), but as yet this has not been presented in peer reviewed literature.

A method which has seen some use in real applications is High Dimensional Model Representaion (HDMR), where 'high dimensional' refers to the input variable space (Li et al., 2001). Model output is again expressed as a function of model inputs, but this time it takes the form of a series expansion of sums of functions, with each sum in the series adding functions of one more input variable than the previous one,

$$f(x) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \le i < j \le n} f_{ij}(x_i, x_j) + \sum_{1 \le i < j < k \le n} f_{ijk}(x_i, x_j, x_k) + \ldots + f_{1\ldots n}(x_1 \ldots x_n) \tag{1.10}$$

where $f(x)$ is a model output and f_0 is its mean value. The first sum in the series is a sum of functions each representing the individual contributions of each of the model inputs, the next is a sum of functions representing the contributions of pairs of inputs, the next sets of three inputs working together, and so on, until the final function in the series represents the combined effects of all of the model inputs. Li et al. claim that in real applications numerical model output is seldom greatly influenced by combinations of more than two input variables, so this reduces to,

$$f(x) \approx f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) \tag{1.11}$$

The authors go on to describe two different procedures for establishing the component functions of the above expansions, both are computationally complex, but the important point again is that it takes fewer model runs to define the HDMR than to perfom a traditional Monte-carlo uncertainty analysis. The method has been applied to various modelled and laboratory studies that have a large number of inputs, such as a human exposure model for trichloroethylene (Wang et al., 2003) and ignition times for gas and air mixtures (Li et al., 2008). Of more relevance here, it was applied to a photochemical box model (Wang et al., 2001) where it was found to run about 400 times faster than the original model, so allowing efficient Monte-carlo analysis. This box model was far simpler than CMAQ, however, and there is little evidence that the assumption that no more than two input variables interact to influence model output is guaranteed to hold for a full Eulerian CTM.

Finally we come onto Kriging as a metamodelling technique, which is essentially an interpolation method, and has its origins in the field of geostatistics, where it is routinely applied over two or three spatial dimensions and takes its name from South African mining engineer D.G. Krige (Cressie, 1993). In the field of computer experiments kriging is commonly referred to as Gaussian process emulation and the resultant metamodel as an 'emulator' (Kennedy and O'Hagan, 2001). The idea of using Gaussian processes to model the output of deterministic computer codes is widely credited to Sacks et al. (1989). It is an ideal method for this purpose as it will pass exactly through all of the points it is trained on, in the same way as a deterministic model would produce those exact same points again if run at the same input settings, even if those training points describe highly non-linear or non-monotonic model output. Another benefit of using

Gaussian process emulation as a metamodelling technique is that the resultant emulator not only provides an estimate of the model output at untried input values (the emulator mean), but also gives the variance of this estimate (Sacks et al., 1989). This idea was developed further in a Bayesian framework, with the aim of incorporating observational data to calibrate the model input factors and estimate model discrepancy, by Kennedy and O'Hagan (2001). The method, the authors claim, can be applied to models of 'arbitrary complexity', but they concede that computational difficulties would arise with high dimensional input spaces. The same statistical framework can in theory be used for fully probabilistic uncertainty analysis (Oakley and O'Hagan, 2002) and sensitivity analysis (Oakley and O'Hagan, 2004). Rougier (2007) explains how this might be applied to climate models in the idealised scenario where the climate modeller is able to make probabilistic specifications of all the required quantities, but Edwards et al. (2010) argue that 'the practical application [of such an analysis] would be extremely challenging, even for relatively simple models'. A 'lack of identifiability' can be a serious problem in this kind of Bayesian calibration (Bayarri et al., 2007), that is, with complex models there can be many different combinations of input variable values which lead to the same value of model output. Kennedy and O'Hagan (2001) suggest that observational data corresponding to model inputs, as well as just the outputs as used in GLUE or BMC, might alleviate this problem. The difficulties posed by lack of identifiability, or 'equifinality', are common to all model calibration methods and has led Beven (2006) to claim that there may be no one best parameter set for a given model, or indeed one best model. This seems counter intuitive as environmental models are intended to simulate a definite physical reality and their inputs real physical quantities, or at least parameterisations derived from real physical quantities. Many authors have been at odds with Bevens' point of view (Stedinger et al., 2008) and this is reflected in a phrase sometimes heard in the UK air quality policy arena, 'right for the right reasons'. It is quite clearly the best case, if at all possible, to have a model which

accurately represents reality, and inputs which do the same. It may be possible, however, to view the position put forward by Beven as considering 'right for the right reasons' to be the ideal scenario, but 'right for the wrong reasons' could be a situation we have to settle for in some practical applications.

Gaussian process emulation is the chosen method for constructing metamodels of CMAQ in this project, and as such further details are given in the next chapter. Also, much of the rest of the literature reviewed here involves Gaussian process emulation in some way or other.

1.7.4 History Matching and Calibration of Climate Models

Returning to the climate modelling example of Williamson et al. (2013), Gaussian process emulation was used in place of the computationally expensive climate model, HadCM3, with the objective of reducing the volume of the input variable space in order to make subsequent, more accurate emulators easier to fit. History matching can achieve this by calculating an implausibility measure for each member of a Monte Carlo sample, using the output of that particular run and corresponding observational data. Those members of the sample with implausibility above a given threshold are then discarded. The mathematical form of this implausibility measure will be defined in the following chapter on the methods used in this project. The observational data used by Williamson et al. (2013) consisted of four separate metrics of highly aggregated average measurements of temperature and precipitation, for which they constructed four separate emulators. For any particular choice of model inputs, the implausibility was calculated using the emulated model output of each metric, and the maximum of these values assigned to that combination of inputs. The history matching process requires the model discrepancy to be specified - to recap, this is the error in model output which would still remain if the correct input values were known and stems from the model's imperfect representation of reality. This can be estimated by gathering expert opinions, but in this case was derived from an ensemble of climate models,

and Williamson et al. describe a framework for doing this. In this framework they state that the discrepancy must be interpreted as the maximum error that the climate modelling community would deem to be acceptable, rather than the actual error for the model in question. Of the four observational metrics, Williamson et al. find that one, global mean surface air temperature, is far more powerful than the other three in terms of the volume of input variable space it is able to rule out. They state amongst other things, that this was due to smaller observational error and the model having a lower discrepancy for this variable. The method used by Williamson et al. for estimating the model discrepancy produced a range of possible values, the nominal value resulting in 56 % of the parameter space being ruled out. In a test of the sensitivity of the history matching process to the highest and lowest values of this range, they found a 14 % difference in the amount of parameter space ruled out.

Again in the climate modelling arena, Edwards et al. (2010), apply history matching to the C-GOLDSTEIN model, and describe it with the term 'precalibration' as an indication of the idea that using it to reduce the parameter space of the model will simplify a following fully Bayesian calibration exercise. They acknowledge that the full calibration may not actually be tractable in many cases, so that precalibration may be seen as doing something pragmatic at the present time, with a view towards future calibration efforts. Key to this idea is the difficulty in specifying model discrepancy; a full calibration exercise, Edwards et al. assert, requires this to be specified as a discrepancy variance matrix, which is a demanding exercise, whereas history matching only requires the discrepancy to be specified as a single variance for each model output against which observations are to be compared. Edwards et al. also calculate the influence of each input factor on the calculated value of implausibility, in an interesting alternative to conventional sensitivity analysis, which would concern the influence of the inputs on the model output directly. They then project the implausibility onto the four most important inputs which allows a graphical interpretation of the results via

a four-way panel of two dimensional plots.

McNeall et al. (2013) suggest an alternative use of the history matching procedure, which is to use it to assess the power of as yet unobtained observational data to constrain model input parameters. Their method involves the use of Gaussian process emulation again, so an ensemble of model runs, a model of the Greenland ice sheet in their example, was made on which to train the emulator. The observational data is replaced by model output from one of the training runs, and the emulator built using the remaining runs. Implausibility is calculated for a Monte Carlo sample of input values in the usual way, except that observational error and model discrepancy are both zero, and this is repeated using each ensemble member in turn as the artificial observational data. In each case the reduction in the range of each input parameter and the reduction in the volume of the input variable space is calculated as a measure of the power of the data to reduce the input variable space if it were real observational data. McNeall et al. found that the potential constraint provided by an observation could be different depending on which ensemble member was used as the observational data, indicating that the usefulness of a particular measurement can vary according to its actual value. They also repeated the experiment with error added to the artificial observations, so demonstrating that it is possible to estimate the level of accuracy required by a potential future measurement campaign.

The above three examples of history matching are all from the area of climate science, as to date this method has not been employed in air quality modelling. The development of the Bayes linear approach to uncertainty analysis, and history matching in particular, were motivated by the difficulties in applying a fully Bayesian methodology to calibrate a model and perform sensitivity and uncertainty analysis. Even with the simplified Bayes linear approach, examples of producing calibrated probabilistic predictions in environmental modelling are rare. Such an analysis was attempted, nonetheless, by Sexton et al. (2011), who make probabilistic projections of climate response to a doubling of CO_2 using the UK

Met. Office HadSM3 model. In common with Williamson et al. (2013), they use an ensemble of climate models to estimate discrepancy. This approach relies on the assumption that the differences between models is comparable to the difference between a single model and the true system, and Sexton et al. concede that this is not ideal, as all of the models may differ from the real system in similar ways. They use a large observational dataset to calibrate the model output, and make this process tractable by reducing the observations to a set of eigenvectors via a singular value decomposition. Sexton et al. do manage to make calibrated predictions of future climate in a scenario of doubled CO_2 emissions, but this involves a number of assumptions and modelling choices, to which the sensitivity of the results is examined in a companion paper (Sexton and Murphy, 2011). These sensitivity tests do appear to show that the probabilistic projections of future climate are reasonably robust to those assumptions which can be tested, but the authors admit that not all assumptions are testable. In particular, they again draw attention again to the use of the multimodel ensemble to produce the discrepancy term, and the fact that this has no way of accounting for systematic errors across all of the models or for unknown physical processes. Nevertheless, Sexton and Murphy should be commended for the honest and open appraisal of their own work.

A number of case studies where Gaussian process emulators have been used for sensitivity and uncertainty analysis are described by Kennedy et al. (2006), following work at the Centre for Terrestrial Carbon Dynamics (CTCD). The first of these was an analysis of the Sheffield Dymanic Global Vegetation Model (SDGVM), in which experts decided on the five most important inputs and assigned ranges to their values. An 80 point Latin hypercube gave design points at which to run the model and provide training data to emulate the single model output of net ecosystem productivity. Soil properties became of interest following the first emulation and a series of nine further emulators were constructed to investigate the effects of other input variables. Another model used at the CTCD is

the Soil Plant Atmosphere (SPA) model which simulates ecosystem photosynthesis and water balance, and was the subject of a sensitivity analysis by Kennedy et al. (2006) of its main output, terrestrial gross primary production. A 150 point Latin hypercube sampling plan was used with nine input variables, and an emulator used to decompose the variance of the output into percentage contributions from each of these nine variables, both individually and in combinations. In order to demonstrate the effectiveness of the emulator, the authors estimated the uncertainty in one of the inputs and propagated it with 5000 Monte-carlo runs of the SPA model. The resulting estimate of terrestrial gross primary production was 3.57 g cm^{-2} day^{-1}, with a variance of 0.05, and the emulator gave an output of 3.56 g cm^{-2} day^{-1} with a variance of 0.06 (Kennedy et al., 2006). The third and final case study from the CTCD was an attempt to estimate the uncertainty in the UK carbon budget using the SDGVM. Kennedy et al. (2006) managed to decompose the variance into contributions from fourteen different inputs, but the estimate of overall uncertainty was not robust because experts were unable to agree on the probability distributions of some of the inputs. These case studies highlight the difficulties in quantifying uncertainty in model output, when the degree of uncertainty in model inputs is itself uncertain.

1.7.5 Formal Sensitivity Analysis in Climate Modelling

The variance decomposition mentioned at the end of the last section is a more formal method of sensitivity analysis than some of the ad-hoc approaches described at the start of this review. It will be explained in more detail in the next chapter, as one of the methods used in this project, but the basic principle is to perturb all of the model inputs at the same time and thereby induce a variance in the output. The amount of this variability which can be attributed to any particular input is quantified using estimates of how much it would be reduced if the true value of that input were known (Saltelli et al., 2008). Various techniques for calculating this are described by Saltelli et al., but all require considerable numbers of model

runs, so the use of emulators provides obvious benefits. One such method, the Fourier Amplitude Sensitivity Test (FAST), which again is elaborated upon in the next chapter, was used by Lee et al. (2012) with a global aerosol model. The modelled concentration of cloud condensation nuclei was emulated with respect to eight uncertain parameters, using ten training runs per parameter. Lee et al. produced a separate emulator for each grid cell in their model, and carried out a separate FAST with each one. This is a pragmatic approach to investigate sensitivity across a spatial domain, as emulation of spatial or spatio-temporal fields is an active research area, but such methods have not been sufficiently validated for use in practical applications (Conti and O'Hagan, 2010). Lee et al. (2013) extended this study to include 28 input variables, again producing emulators for each grid cell of their global model domain, and this time using the emulators for Monte Carlo uncertainty estimation as well as sensitivity analysis. This time they used only six training runs per variable, citing tests carried out during the construction of emulators for their previous study (Lee et al., 2012), which in contrast involved only eight input variables. The authors performed a further 84 runs for the purpose of emulator validation, but this validation concentrates on showing that when the emulators were used to predict the output of the validation runs, the correct proportion of those validation runs lay within a particular emulator confidence interval. This is fine if the both the emulator mean and variance are to be used in subsequent analyses, but the results presented in the paper use only the emulator mean predictions, so it is difficult to see how such a validation is relevant. It would be more informative to present some summary statistics of the emulators skill in predicting the output of such a substantial validation sample.

1.8 Structure of the Thesis

As is evident from the preceding review of literature, the only uncertainty analyses of the CMAQ model have been for US domains and have made use of partially validated RFMs extrapolated from local sensitivity coefficients. Also, the ad-hoc

one-input-at-a-time type sensitivity analyses which have been performed, whilst useful to answer specific scientific questions, cannot uncover the pattern of sensitivity to input uncertainties revealed by a full global analysis. These issues have not previously been addressed due to the difficulties involved in performing enough CMAQ runs for robust analysis. The results presented in this thesis aim to demonstrate that uncertainty and sensitivity analysis of the complex CTMs used in air quality modelling can be made feasible using Gaussian process emulation techniques, and that valuable insights into such models and the underlying systems which they represent can be gained in this way. The first variance-based global sensitivity analysis of CMAQ has been performed, which addresses the issue of which input uncertainties dominate the modelled ozone output uncertainty for a summertime photochemical pollution episode in the UK.

The methods used to produce the results given in later chapters will be explained in detail in chapter 2, and chapter 3 describes the screening of input factors to find a subset which have the most influence on modelled ozone and NO_2 concentrations, to be carried forward for further analysis. This includes the use of Morris' method, which we shall encounter for the first time in chapter 2 as it does not appear in the literature on the sensitivity analysis of complex environmental models. In chapter 4 the results of variance-based sensitivity analyses will be presented, making use of emulators in place of the CMAQ model, to show how the sensitivity of model output to uncertainty in different input factors changed over the period of the July 2006 ozone pollution episode. Also in this chapter, the effect of varying single inputs, such as NO_x emissions, on peak ozone values will be examined, whilst taking into account the uncertainty in other inputs. The final section of chapter 4 uses bivariate plots of peak ozone response to changes in NO_x and VOC emissions to address the policy relevant question of which emissions reduction strategy would have the greatest positive effect on ozone concentrations. The use of the same emulators to allow efficient Monte Carlo uncertainty propagation will be described in chapter 5, along with

some initial forays into BMC and history matching. Finally, chapter 6 will sum up the novel aspects of the results which have been presented and will provide some concluding remarks and suggestions for further work.

The results of some of the sensitivity analyses described in chapter 4 have been presented as a poster at the 2014 Uncertainty in Computer Modelling conference at the University of Sheffield. The same results are in preparation for publication in the journal *Environmental Science and Technology*, and an abstract has been accepted for oral presentation at the 2015 Community Modelling and Analysis System (CMAS) conference at the University of North Carolina.

Chapter 2

Methods

This chapter describes the methods used in the sensitivity and uncertainty analysis of the CMAQ model. As mentioned previously, this traditionally requires a large number of model runs, and therefore the methods chosen are those which will extract the maximum amount of information from the smallest number of runs. Section 2.1 describes Morris' method for screening the large number of uncertain input variables in order to identify those which have the greatest influence on model output uncertainty, and should therefore be carried forward into further analyses. Despite the reduction in the number of input variables following the screening process, the remaining analyses still could not be carried out directly on the CMAQ model. To circumvent this problem, Gaussian process emulators were used in place of the model, and this procedure is described in section 2.2. Sensitivity analyses were performed using the extended Fourier Amplitude Sensitivity Test (FAST) method of Saltelli et al. (1999), which is described in section 2.3. Observational data is brought into the analysis with Bayesian Monte Carlo in section 2.4, and history matching in section 2.5. The observational data used is described in section 2.6.

Even with the benefits of emulation, a far greater number of model runs were required than could possibly have been carried out manually, so section 2.7 describes an automated system which was created to facilitate this process.

2.1 Input Variable Screening with Morris' Method

It is common for large computer models with many input variables to have only a small subset of inputs which significantly influence a particular output (Saltelli et al., 2008). In view of this, a useful exercise is to screen the inputs in order to identify this subset, hence significantly reducing the computational burden of subsequent sensitivity and uncertainty analyses.

The method chosen here for this purpose is that described by Morris (1991), which has the benefits of being easy to understand and implement, whilst at the same time giving an indication of variable interaction and model non-linearity using a manageable number of model runs. It is often referred to as Morris one-at-a-time (OAT) method, which is slightly misleading, as although each input is indeed altered one at a time whilst holding the remainder of the inputs constant, this is done several times for each input variable, with the remainder in different positions on each occasion. Furthermore, each of these perturbations of the variable in question is made over a different portion of its entire range of variability. In this way, Morris' method bridges the gap between what are termed 'local' and 'global' sensitivity analysis techniques. A local sensitivity analysis is one in which the input factors are varied by a small amount around their nominal values, whereas in a global sensitivity analysis they are perturbed over their full ranges of variability, hence exploring the entire input variable space (Saltelli et al., 2008).

Morris defines a k-dimensional grid with p levels over the space of the k input variables, $x_1 \dots x_k$, where each x_i may take values from $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$. If Δ is a multiple of $1/(p-1)$, Morris calculates the 'elementary effect', d_i of input x_i as follows,

$$d_i(\mathbf{x}) = \frac{[y(x_1, x_2, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k) - y(\mathbf{x})]}{\Delta} \tag{2.1}$$

where $x_i \leq 1 - \Delta$, so that the design does not go outside of the input variable space, and $\mathbf{x}$ is any point on the grid defined in that space. The strategy described by Morris to evaluate the relative importance of the input factors is to calculate a number of these elementary effects and produce summary statistics of their distributions. In order to calculate one elementary effect, two model runs are required, so if r elementary effects are desired for each factor, then the total number of model runs required will be $2rk$. Morris devised a simple yet effective sampling scheme to reduce this number and improve the efficiency of the method. A point on the grid is chosen at random and then one x_i is increased by Δ, generating a new point. From this point a different x_i is increased by Δ, generating a third point and so on until all k factors have been incremented. The result is a trajectory defined by $k + 1$ points which moves around the input space in steps parallel to its axes, and which allows one elementary effect to be calculated for each factor. This means that if r elementary effects are to be calculated for each of k factors, the total number of model runs required is $r(k + 1)$. An example is shown in figure 2.1 where five trajectories have been created in a two dimensional input space, so here a total of fifteen model runs are used to produce ten elementary effects.

After calculating the r elementary effects for each input factor, Morris recommends calculating their means, μ, and standard deviations, σ. A high value of μ for a particular input factor indicates that the factor has an important effect on the model output. If all of the elementary effects for the factor are similar, then σ will be low, indicating that the factor has a fairly linear effect on the model output and is not greatly involved in interactions with other factors. If, however, σ is large then this shows that the elementary effects for the factor varied widely, indicating either that its effect on model output is non-linear, or that it is involved in interactions with other factors. A model with non-monotonic output will produce some positive and some negative elementary effects, giving a small value for μ and a high σ, meaning that the two quantities must be considered

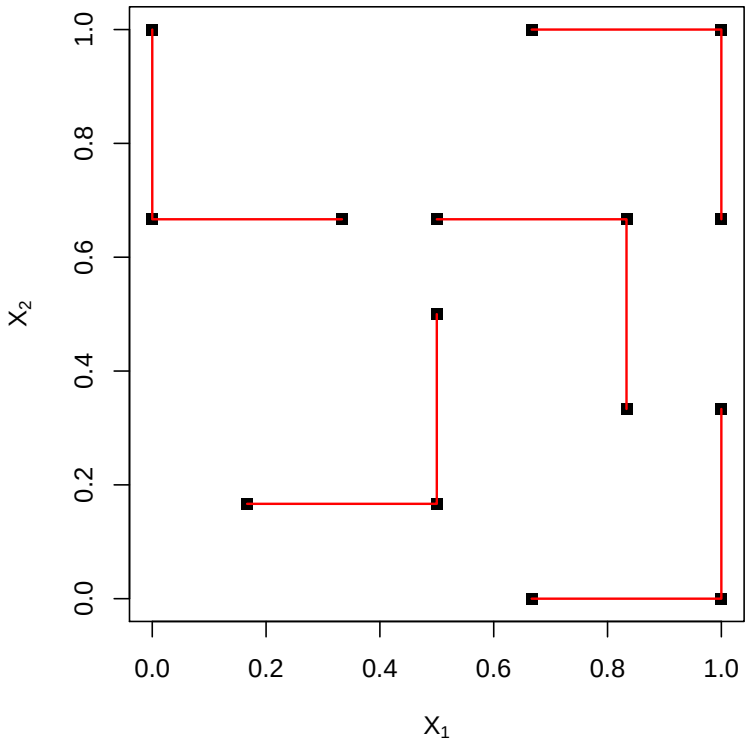


Figure 2.1: Five Morris' method trajectories in a two dimensional input space

together in order to rank the input factors.

A useful addition to the method, suggested by Campolongo et al. (2007), is to use the mean of the absolute values of the elementary effects, denoted by μ^*. This provides a single measure which can be used to rank the input factors in order of importance, even for non-monotonic models. In the same paper Campolongo et al. (2007) suggest a further improvement to the method, in order to improve the space filling properties of a collection of trajectories. A large number of trajectories is generated, with the intention of keeping a subset which have the largest combined distance between them. The sum of the distances between all the pairs of points in a pair of trajectories is calculated, and then summed over all possible pairs of trajectories in a randomly chosen subset of the required size. When this has been completed for all possible subsets of trajectories, the one with the largest total distance is retained. This space filling method is implemented in the function `morris` of the package *sensitivity* (Pujol et al., 2014) in the R statistical computing language (R Core Team, 2014), and was used in all of the

screening exercises described in chapter 3.

2.2 Gaussian Process Emulation

This section is intended to provide a conceptual understanding of Gaussian process emulation in order to facilitate its practical application using standard tools, although some mathematical details are provided for the sake of completeness. As mentioned in the previous chapter, the method has its roots in the field of geostatistics, where it is generally termed Kriging, but when generalized to higher dimensions can be used as a flexible non-linear regression technique. It is perhaps helpful to the non-statistician to point out that an emulator is simply a regression model of the original computer model, with the computer model output as the dependent variable and a range of values of its inputs as the independent variables. This means that if a number of runs are made at points distributed across the model's input variable space to provide training data with which to build the emulator, it can then be used to predict the output at other input values within the range of the training data.

In a one dimensional example, imagining the model output as a function of a single input variable, the value of the function evaluated at any particular input value can be thought of as a single sample from a Gaussian distribution. If a number of values of that input variable are chosen, then we have a number of samples each from its own Gaussian distribution. Extending this idea to the infinite number of points described by a continuous function, gives an infinite number of Gaussian distributions, known as a Gaussian process. More generally, in statistical terminology, a stochastic process is a collection of random variables which describe the variation in some quantity over time or space, with one random variable representing each point in that space at which the quantity is evaluated. In the case of Gaussian process emulation of computer models, the quantity which is varying is the model output and the space over which it varies is the input variable space. As the CMAQ input variables being considered are all continuous,

we have, at least in theory, an infinite number of points within the space they delineate, and hence an infinite number of random variables representing the output, which can now be thought of as a random function. There are a number of ways of conceptualising Gaussian processes, but this 'function-space view', as Rasmussen and Williams (2006) refer to it, is perhaps the easiest to grasp for those encountering Gaussian process emulation for the first time. Rasmussen and Williams (2006) go on to explain that the properties of this function at a finite number of points, which are found by inference from a finite subset of distributions, will be the same as if the infinite number of distributions had been considered. In other words, the properties of the random function representing the computer model output, which are inferred from the training runs, will be the same across the whole input variable space, allowing inference about the value of the model output to be made for values of the input variables at which the model has not been run.

Taking the usual notation where the computer model output, $y(\cdot)$, is a function of its inputs, $\mathbf{x} = (x_1 \dots x_k)$, then Sacks et al. (1989) treat it as the sum of a regression model and a random process,

$$Y(\mathbf{x}) = \sum_{i=1}^{k} \beta_i f_i(\mathbf{x}) + Z(\mathbf{x}) \tag{2.2}$$

where the random (in this case Gaussian) process, $Z(\mathbf{x})$, is the mechanism that transforms the deterministic function, $y(\mathbf{x})$, into the random function, $Y(\mathbf{x})$. Predictions of the computer model output at points outside of the training data are given by the Kriging mean, which is essentially a weighted sum of the output values of the training data, and each prediction has an associated Kriging variance which gives an indication of the reliability of the prediction. Quite naturally then, the optimum values of the weights are those which minimise the Kriging variance. Following the notation of Roustant et al. (2012), we denote the input points of

the training data (CMAQ runs) by $\mathbf{X} = (\mathbf{x}^1 \dots \mathbf{x}^n)$ and the corresponding output by $\mathbf{y} = (y(\mathbf{x}^1) \dots y(\mathbf{x}^n))^\top$. In the case where the regression function is known, prediction at new points is termed 'simple Kriging', and simplifying the notation by letting $\mu(\mathbf{x}) = \sum_{i=1}^{k} \beta_i f_i(\mathbf{x})$, Roustant et al. give the following expression for the Kriging mean at any given $\mathbf{x}$,

$$m_{SK}(\mathbf{x}) = \mu(\mathbf{x}) + \mathbf{c}(\mathbf{x})^\top \mathbf{C}^{-1}(\mathbf{y} - \mu(\mathbf{X})) \tag{2.3}$$

where $\mathbf{c} = (C(\mathbf{x}, \mathbf{x}^i))_{1 \leq i \leq n}$ is the vector of covariances between $Y(\mathbf{x})$ and $Y(\mathbf{X})$, and $\mathbf{C} = (C(\mathbf{x}^i, \mathbf{x}^j))_{1 \leq i,j \leq n}$ is the covariance matrix of $Y(\mathbf{X})$, the Kriging weights being given by the product $\mathbf{c}(\mathbf{x})^\top \mathbf{C}^{-1}$. The covariance function, C, relies on the assumption that the model output is a smooth, continuous function of its inputs, which implies that its value will change by a smaller amount between points which are nearby in the input space than between points which are further apart in that space. This covariance function takes the general form,

$$C(\mathbf{x}^i, \mathbf{x}^j) = \sigma^2 R(\mathbf{x}^i - \mathbf{x}^j) \tag{2.4}$$

where R is a correlation function which depends on the distance between the two points in the input space, and σ^2 is the 'process variance' (Roustant et al., 2012). $R(\cdot)$ decreases as the distance between the two points increases and also satisfies the condition $R(\mathbf{x}^i - \mathbf{x}^i) = 1$ (Oakley and O'Hagan, 2002). The Kriging variance is then given by,

$$s^2_{SK}(\mathbf{x}) = C(\mathbf{x}, \mathbf{x}) - \mathbf{c}(\mathbf{x})^\top \mathbf{C}^{-1} \mathbf{c}(\mathbf{x}) \tag{2.5}$$

The case where the coefficients, $\boldsymbol{\beta} = (\beta_1 \dots \beta_k)$, of $\mu(\mathbf{x})$ are unknown and

must be estimated is known as universal Kriging, and the mean and variance are given by (Roustant et al., 2012; Oakley and O'Hagan, 2002),

$$m_{UK}(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \widehat{\boldsymbol{\beta}} + \mathbf{c}(\mathbf{x})^\top \mathbf{C}^{-1}(\mathbf{y} - \mathbf{F}\widehat{\boldsymbol{\beta}}) \tag{2.6}$$

and

$$s^2_{UK}(\mathbf{x}) = s^2_{SK}(\mathbf{x}) + (\mathbf{f}(\mathbf{x})^\top - \mathbf{c}(\mathbf{x})^\top \mathbf{C}^{-1}\mathbf{F})^\top (\mathbf{F}^\top \mathbf{C}^{-1}\mathbf{F})^{-1}(\mathbf{f}(\mathbf{x})^\top - \mathbf{c}(\mathbf{x})^\top \mathbf{C}^{-1}\mathbf{F}) \tag{2.7}$$

where $\mathbf{f}(\mathbf{x})$ is the vector of the values of the regression functions at $\mathbf{x}$ and $\mathbf{F} = (\mathbf{f}(\mathbf{x}^1) \dots \mathbf{f}(\mathbf{x}^n))^\top$ is the matrix of such values given by the training runs. $\widehat{\boldsymbol{\beta}}$ is the best linear estimator of $\boldsymbol{\beta}$ and is given by $\widehat{\boldsymbol{\beta}} = (\mathbf{F}^\top \mathbf{C}^{-1}\mathbf{F})^{-1}\mathbf{F}^\top \mathbf{C}^{-1}\mathbf{y}$.

2.2.1 R Packages for Emulation

This section describes the construction of Gaussian process emulators using user contributed R packages, in a way which is accessible to the deterministic modelling practitioner. The use of R for constructing emulators is attractive because of the language's convenient data manipulation features, particularly functions which are able to read from netCDF files, which CMAQ uses as its main input/output file format. Two packages were tested side by side in an 'out of the box' fashion, with parameters kept at default values. Specifically, the *BACCO* (Hankin, 2005) and *DiceKriging* (Roustant et al., 2012) packages are compared.

The data used to perform this test were taken from a sensitivity analysis experiment performed while testing the methods and developing the automation routines used in the project. 66 variables from the emissions and boundary conditions model inputs were screened using Morris' method with ten trajectories, requiring a total of 670 CMAQ runs. This reduced the number of variables to a shortlist of 16, and a further set of 336 CMAQ runs were performed so that em-

ulators could be constructed using 21 runs per variable, a higher number than is commonly seen in the literature (for example Lee et al. (2012, 2013)). The reason for choosing such a high number was to allow emulators with a small variance to be constructed, so that the emulator mean could be used alone in subsequent analyses, where appropriate, greatly reducing their complexity as compared to using both the mean and the variance. More generally, any emulator should be as accurate as possible in order to promote confidence in results which are obtained during its use, so a large number of CMAQ runs were performed simply because the capability to do so had been developed. In the test presented here, 36 of the model runs were selected at random and held back from the emulator construction, leaving 300 runs to build the emulator, amounting to 18.75 runs per variable. The emulators constructed with both packages were used to predict the model output ozone concentrations from the 36 out of sample runs, and this is shown plotted against the real model output in figure 2.2. *DiceKriging* clearly outperforms *BACCO* in this test, and on this basis, combined with the fact that it is considerably faster, was chosen for use in the rest of the project.

The *DiceKriging* package, with its default settings, implements universal Kriging with the Matern(5/2) covariance function. If the distance between two values of the same input variable is h, then,

$$c(h) = \left(1 + \frac{\sqrt{5}|h|}{\theta} + \frac{5h^2}{3\theta^2}\right) \exp\left(-\frac{\sqrt{5}|h|}{\theta}\right) \tag{2.8}$$

where θ is a 'length-scale' or 'range' parameter and controls the distance from a point to be predicted at which training points will cease to have an influence on that prediction and hence the overall smoothness of the interpolation. The default (and only) covariance function in the *BACCO* package is an exponential covariance function, and these different functions are likely to be one of the main causes of the difference in performance between the two packages.

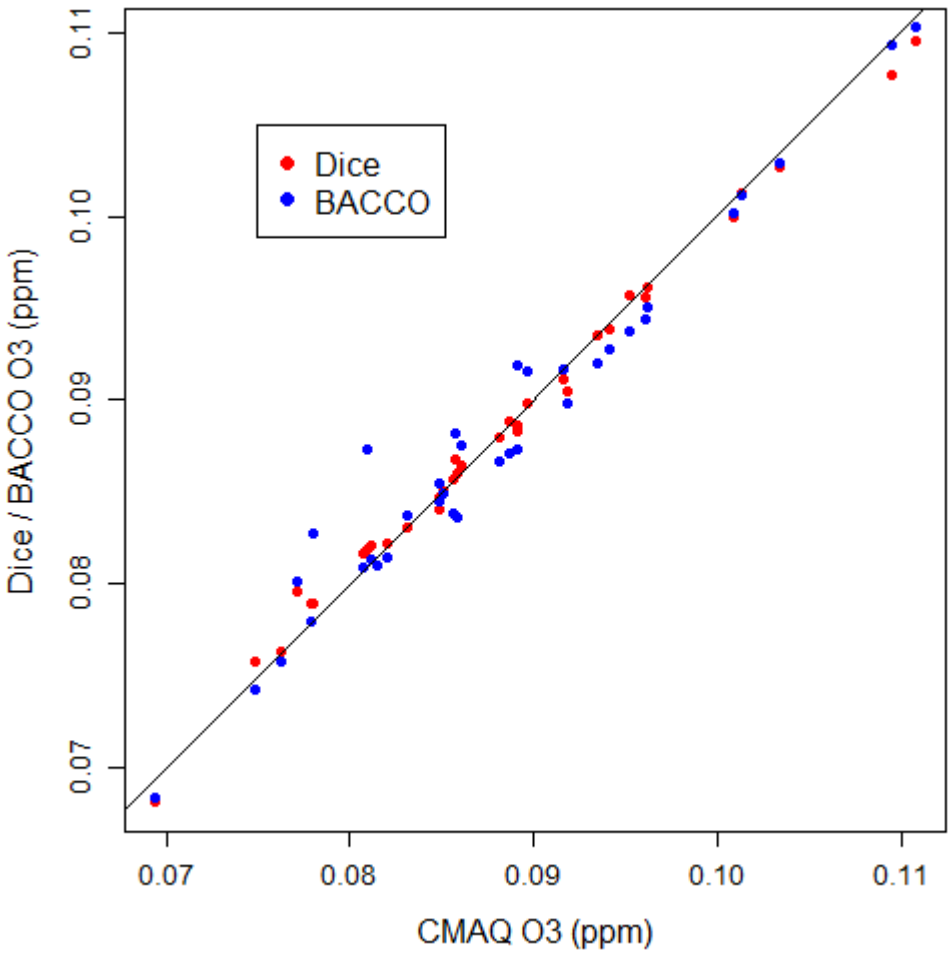


Figure 2.2: 36 out of sample points, as predicted by emulators trained on 300 points, by the *BACCO* and *DiceKriging* packages. Predictions from a perfect emulator would lie on the black 1:1 line.

2.2.2 Latin Hypercube Designs

This section is concerned with the method of choosing which values of the input variables to run the deterministic computer model at in order to provide training data to build an emulator. What is required, essentially, is a method of positioning a predetermined number of points in a high dimensional space in a way which will fill that space as effectively as possible. The method chosen to accomplish this task is Latin hypercube sampling (McKay et al., 1979), which is a form of stratified sampling. The hypercube is a generalisation of a cube into a higher dimensional space, specifically if there are k input factors we will have a k-dimensional hypercube with axes $X_{1,\ldots,k}$. McKay et al. designed this method

as a more efficient alternative to simple Monte Carlo sampling for the uncertainty analysis of computer models with known probability distributions on the inputs. If there are to be n model runs then the range of each input is divided into n intervals of equal probability. When using the method to choose design points for building an emulator, the factors are given uniform distributions resulting in n equally spaced intervals on each axis, and then one sample of each factor is made in each interval. One of the samples from X_1 is selected at random and matched with a sample selected at random from X_2,X_3 up to X_k, giving the first point, $\mathbf{x}_1$ of the Latin hypercube sample. This sampling continues without replacement until all of the intervals have been used up and the final point $\mathbf{x}_n$ has been chosen. An example is shown in figure 2.3 with 20 points and 2 dimensions, hence this special case could be called a Latin square. Note how the points cover the range of both axes evenly.

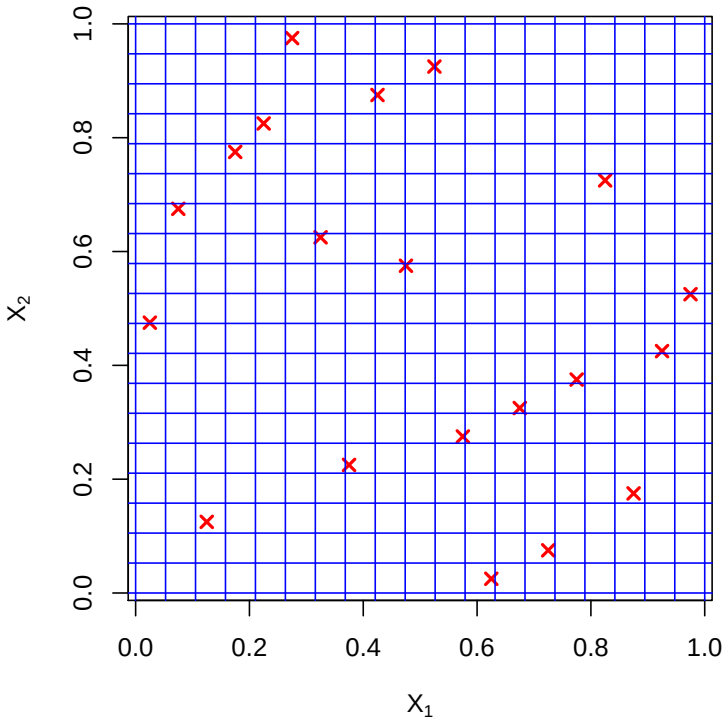


Figure 2.3: A 20 point 2 dimensional Latin hypercube sample with uniform distributions on each dimension

2.3 Sensitivity Analysis

The concept of local and global sensitivity analysis was introduced in section 2.1, and it is on global sensitivity analysis that this section focuses - examining the effect that uncertainty in the values of model input variables has on model output over their full ranges of variability. Also important in a global analysis is that all of the inputs are perturbed at the same time, so that the sensitivity to a particular input is not assessed in isolation from uncertainty in other inputs. The metric of choice for this sensitivity is that proposed by Sobol' (1990) which is based on decomposing the variance in model output caused by input uncertainty, and is robust in the face of model non-linearity and non-monotonicity.

2.3.1 Sensitivity Indices

The idea employed by Sobol' (1990) was to decompose the output variance into a sum of terms of increasing dimensionality as follows,

$$V_y = \sum_{i=1}^{k} V_i + \sum_{1 \leq i < j \leq k} V_{ij} + \ldots + V_{1,2,\ldots,k} \tag{2.9}$$

where each of the terms in the expansion give the variance attributable to inputs x_1, $x_{1,2,\ldots,k}$ and so on. The sensitivity indices are then given by dividing each of the terms in this decomposition by the total variance, so that the first k terms give the first-order effects for each input, then the terms involving two inputs give the second-order effects and so on. In this way the sensitivity indices naturally represent the fraction of model output variance attributable to each combination of inputs.

The first-order effect, or main effect, of an input factor can be thought of as the amount by which the model output variance would be reduced if the value of that factor were to be fixed. Generally we do not know what value this factor should be fixed at (otherwise there would be no uncertainty in the first place!),

so this reduction in variance must be averaged over all of the possible values of that input. This would give the quantity,

$$E_j(V_{-j}(Y|X_j)) \tag{2.10}$$

for input j, where the expectation is taken over j and the variance is taken over all other inputs except j. The more important input j is, then the smaller the expectation 2.10 will become. Now because,

$$V_y = E_j(V_{-j}(Y|X_j)) + V_j(E_{-j}(Y|X_j)) \tag{2.11}$$

the quantity $V_j(E_{-j}(Y|X_j))$ becomes larger as input j becomes more important. This allows the definition of the main effect of j as (Saltelli et al., 2010),

$$S_j = \frac{V_j(E_{-j}(Y|X_j))}{V_y} \tag{2.12}$$

and in the same way the combined effect of two factors, i and j, would be calculated as,

$$S_{ij} = \frac{V_{i,j}(E_{-i,j}(Y|X_i, X_j))}{V_y} \tag{2.13}$$

All higher-order effects can be calculated by conditioning the expectation in the numerator of 2.13 on increasing numbers of factors. The total number of terms in the variance decomposition for a model with k factors is $2^k - 1$ and hence becomes prohibitive to calculate. Instead, a total effect index may be used, which for a particular factor is the sum of all the terms in which involve that factor, and

would be given by (Saltelli et al., 2010),

$$S_{Tj} = \frac{E_{-j}(V_j(Y|\mathbf{X}_{-j}))}{V_y} \tag{2.14}$$

where the bold face $\mathbf{X}$ represents a vector of random variables. This arises because the main effect of $\mathbf{X}_{-j}$ is given by $V_{-j}(E_j(Y|\mathbf{X}_{-j}))$, so using identity 2.11 again and dividing through by V_y, one arrives at,

$$\frac{E_{-j}(V_j(Y|\mathbf{X}_{-j}))}{V_y} = 1 - \frac{V_{-j}(E_j(Y|\mathbf{X}_{-j}))}{V_y} \tag{2.15}$$

so that the total effect of input j is found by subtracting the main effect of everything that does not involve j from the total variance.

In order for the indices described above to correctly represent the proportion of output variance which is attributable to each factor, it is a necessary condition that those factors are independent. There is a high likelihood that some of the CMAQ input factors are not independent, but there is no evidence of a correlation structure between the uncertainties in those factors. All of the variance-based sensitivity indices presented in this report should be considered only as sensitivities to uncertainty in the inputs and not be interpreted in any other way. While an analysis which takes correlations between the inputs into account is theoretically possible, the extra complexity involved means that a first sensitivity analysis should always be performed with the assumption of orthogonal inputs (Saltelli et al., 2004).

The calculation of the main effects and total effects requires the evaluation of high dimensional integrals, which must be estimated numerically (Saltelli et al., 2008). It is one of the methods to accomplish this to which we turn in the next section.

2.3.2 FAST

The Fourier Amplitude Sensitivity Test (FAST) was originally developed in the 1970s (Cukier et al., 1978) and provides a computationally efficient procedure for estimating the main effects of equation 2.12. The method relates the probability distribution representing the uncertainty in each input factor to a different frequency. The input variable space of the model is then explored by a search curve which traverses each dimension at a rate which is proportional to the frequency assigned to that dimension. This curve defines a set of values of the input factors at which to run the model, which will induce a periodicity in the output. Fourier analysis of this output then gives Fourier coefficients from which the main effect for each factor can be derived. For the reader not familiar with Fourier analysis, the central idea is that a complex periodic signal can be represented as the weighted sum of a set of simpler periodic signals, namely sines and cosines, with the weights taking the form of the aforementioned Fourier coefficients.

The search curve is translated from the n-dimensional space of the input variables $x_1, \ldots, x_n$ into a one-dimensional space parametrised by the variable s using a set of transformation equations,

$$x_i = G_i(\sin \omega_i s), \qquad i = 1, \ldots, n \tag{2.16}$$

where ω_i is the frequency assigned to that particular x_i and the function G_i is chosen so that the fraction of the length of the search curve which lies between x_i and $x_i + dx_i$ is equal to the probability that the true value of the input lies in the same range, according to the probability distribution assigned to describe its uncertainty (Cukier et al., 1978). Cukier et al. go on to describe that if the set of frequencies are chosen so as to be incommensurate, that is their weighted sum cannot be equal to zero, and s varies in the range $\pm\infty$ then the search curve will be open ended and will eventually cover the whole of the input variable

space. However, as computers can only store floating point numbers to a certain precision, it is not actually possible to choose a set of incommensurate frequencies, but the authors develop the theory with integer frequencies and s varying between 0 and 2π, which gives a closed search curve, and then show that the errors this approximation produces are negligible.

The extended FAST method, developed by Saltelli et al. (1999), improves on the original FAST by introducing an improved method for defining the path of the search curve, and by estimating both the main and total effects indices at the same time. This is achieved by choosing two frequencies for each input factor, one of which is assigned to that factor, as in the classic FAST, and the other is assigned to all the remaining factors. This induces two periodicities in the model output for each input, one of which can be used to calculate the main effect, and the other to calculate the total effect of the factor. This extended FAST is implemented in the function `fast99` of the R package *sensitivity* (Pujol et al., 2014), the same package that was mentioned earlier in the discussion of Morris' method.

2.4 Bayesian Monte Carlo

Bayesian Monte Carlo (BMC), as introduced in chapter 1, involves taking a standard Monte Carlo sample of model runs (or emulator 'runs') and calculating weights for each of them according to their proximity to observational data. The mechanism for doing this is Bayes theorem, which takes the general form (Sivia and Skilling, 2006),

$$\Pr(\theta|D) = \frac{\Pr(D|\theta)\Pr(\theta)}{\Pr(D)} \tag{2.17}$$

where $\Pr(\theta)$ is the prior probability of a (statistical) model parameter having a particular value, and $\Pr(D|\theta)$ is the probability that a particular data value

would be observed, given the parameter value, which is termed the likelihood. When these quantities are multiplied together and divided by $\Pr(D)$, which acts as a normalizing constant to ensure all probabilities sum to one, the posterior probability density, $\Pr(\theta|D)$, of the parameter given the observed data is obtained.

BMC takes advantage of the fact that in a conventional Monte Carlo analysis all model runs are assumed to be equally likely representations of reality, so for n model runs the prior probability for each is simply $1/n$. If the observational error can be assumed to be normally distributed then the likelihood of each model run being correct can be calculated using a likelihood function based on the normal probability density function (Dilks et al., 1992),

$$L(O|y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{O-y}{\sigma}\right)^2\right) \tag{2.18}$$

where O is the observational datum, y is the model output and σ is the observational error. In a practical application, once the likelihoods have been calculated for all of the model runs in the Monte Carlo sample, the prior probabilities are simply incorporated into the normalizing constant as each likelihood is divided by the sum of all the likelihoods.

The model output given by the original Monte Carlo sample can be plotted as a probability distribution to represent the uncertainty around the single valued deterministic model output. After weighting the members of that sample by their posterior probabilities an updated distribution can be plotted which is effectively a synthesis of the model output and the observational data. The same process also updates the uncertainty distributions of the model inputs, thereby calibrating both the inputs and outputs at the same time.

To those familiar with the normal density function, it may seem odd to use the observational error for σ, when it would be more conventional to use the standard deviation of the errors. However, all of the applications of BMC in the literature

reviewed in chapter 1 use the observational error. This will be discussed further when the results are presented.

2.5 History Matching

History matching also entails emulating a number of model runs, but here uniform distributions are usually assigned to the input factors, as the aim is to find implausible parts of the input variable space, rather than to produce calibrated output. An implausibility measure, I, is calculated for each model run, $f(x)$, as follows,

$$I^2(x) = \frac{(E(f(x)) - z)^2}{var(f(x)) + var(\epsilon_{mod}) + var(\epsilon_{obs})} \tag{2.19}$$

where $E(f(x))$ and $var(f(x))$ are the emulator mean and variance and z is the corresponding observational datum. ϵ_{mod} is the model discrepancy, i.e. the difference between the real system and the model output if the true x were known, and ϵ_{obs} is the observational error. This equation is taken from Vernon et al. (2010), and has a different form to that shown by Williamson et al. (2013), but is directly equivalent, and is the chosen form here because it is easier to see the influence of the variances in the denominator on the implausibility of a particular set of inputs. Decreases in the accuracy of emulation, shown by an increase in $var(f(x))$, and increases in model discrepancy or observational error will all serve to decrease the implausibility, and therefore reduce the amount of the input variable space which can be ruled out. In order to actually rule out regions of input space a threshold for I must be chosen, and Vernon et al. (2010) recommend a value of 3, citing the 3σ rule, which states that the probability of any unimodally distributed random variable being within three standard deviations of its mean is at least 95 % (Pukelsheim, 1994).

2.6 Observational Data

Both BMC and history matching require the use of observational data, which were taken from the on-line archive of the Automatic Urban and Rural Network (AURN) (DEFRA, 2014). This is a network of fixed, ground-based monitoring stations throughout the UK which is used for assessing compliance with air quality directives. Ozone is measured using UV absorption instruments, which are required to be within $\pm$ 5 % of a reference standard instrument in order for the data to be ratified (AEA, 2007). As measured ozone concentrations in the UK are generally below 100 ppb, and no information is available to indicate bias in the measurements, the observational error was assigned a normal distribution with a standard deviation of 2.5 ppb. Although the AURN measures a number of pollutants, just ozone measurements were used in the present work, which represents initial steps in CMAQ calibration. An issue which is of concern when comparing Eulerian model output with monitoring data is 'change of support' (Berrocal et al., 2010) - that is, the measured data represents a single point in space whereas the model output represents the average concentration across a grid cell, in this case 9 km $\times$ 9 km, with a height of around 15 m above ground level. In order to circumvent this issue measured data was used from the AURN site at Harwell in Oxfordshire, which is classified as a rural background location, and as such is sited so as to be representative of an area at least as large as the CMAQ grid cell size. The area surrounding this site is shown in figure 2.4, where the monitoring station itself is shown by the red cross, and blue grid lines are 1 km apart. This is typical of a rural location in the UK - or in England at least - being surrounded by open land, but at the same time not far from small towns and villages and the roads which connect them.

The measured data from Harwell for 11th - 31st July 2006 is shown in figure 2.5, along with the output from the base case model run. Overall the model seems to capture the observed behaviour quite well, but fails to reproduce some of the afternoon peaks. This is particularly noticeable when concentrations are

highest on the 18th and 19th. This was one of the main reasons for choosing this particular time period - the sensitivity of model output to input uncertainties is most interesting when an important pollution episode occurs which the model only partially captures.

Figure 2.4: Detailed surroundings of Harwell, Oxfordshire. The location of the monitoring station is shown by the red star in the centre of the map. ©Crown Copyright and Database Right [2015]. Ordnance Survey (Digimap Licence).

Corresponding meteorological observations used were recorded at the UK Met. office land surface station at Benson, which is approximately 10 km east of Harwell. Data is stored in MIDAS (Met Office Integrated Data Archive System) and was accessed online via the Centre for Environmental Data Archival (CEDA, 2015).

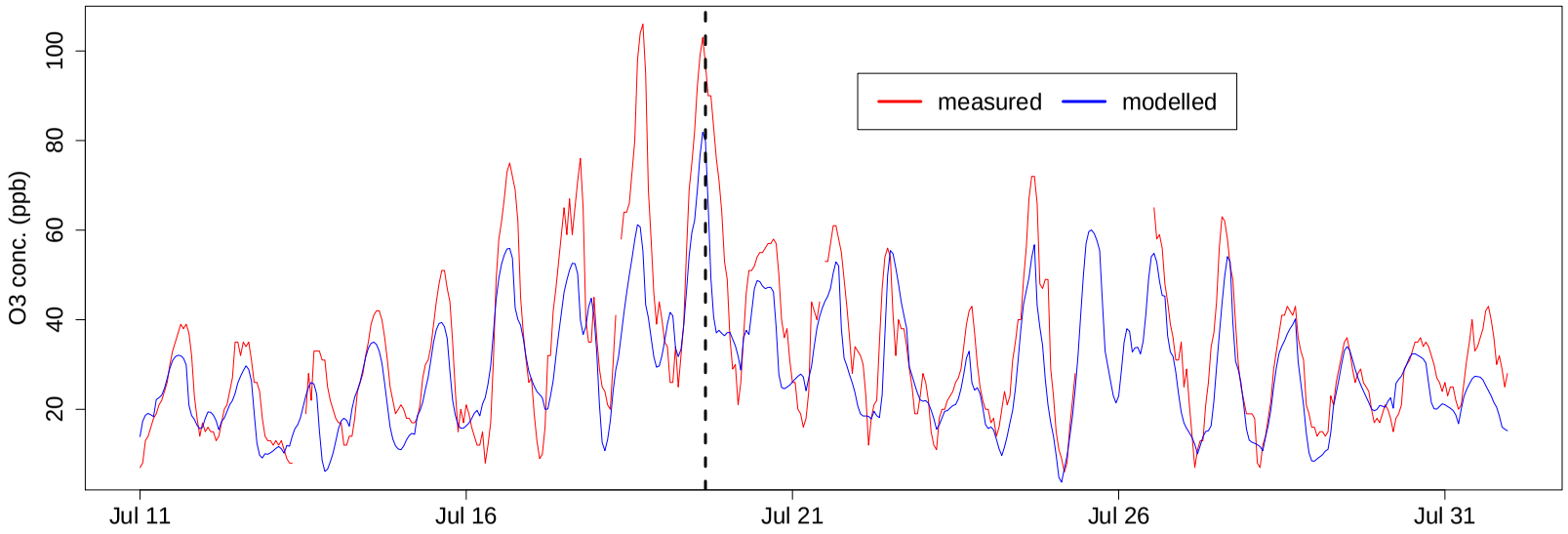


Figure 2.5: measured (red) and modelled (blue) ozone concentrations at Harwell, Oxfordshire, 11th - 31st July 2006. The dashed black line indicates 4 pm on the 19th, a point in time mentioned frequently in later chapters.

2.7 Model Automation

The analysis described in this document began by perturbing emissions and boundary conditions inputs to the model, as these were the most straightforward input variables to modify, being supplied as values in input files. Once a system to do this and to perform large numbers of model runs in an automated fashion had been produced and was running successfully, the scope of the project became more ambitious, and chemical reaction rates and ozone deposition velocity were added to the analysis. These variables were more difficult to modify, as they are hard-coded into the model executable, so the source code must be modified and recompiled for every model run. As such the system described here was actually built in a piecemeal fashion as the project progressed. For the sake of clarity, it will be explained in a qualitative manner, with code listings confined to the appendices.

Another consideration when working with computationally intensive models is the number of computing cluster processors available and how best to use them. Processes which are compute intensive benefit from parallelisation, splitting the task between many processors. However processes which are read/write intensive generally do not, as most computing clusters have all processors writing to the same disk, and only one can access it at a time. The system described herein consists of both types of processes, with on-the-fly creation of input files being read/write intensive, while running the model itself is compute intensive. Despite this, compiling the model to run in parallel would not make the best use of available resources when many model runs must be made, as there is a diminishing return in terms of improving run times as more and more processors are used. This meant that compiling the model to run on a single processor and using the available computing power to execute many runs at the same time was the most efficient use of resources, but introduced its own problems in terms of the volume of input and output data required. The analyses described in the following chapters typically required between three and eight hundred model runs to be

performed, so these would be split into sets of the same size as the number of available processors. For example, if 48 processors were available, and 576 model runs were to be made, this would be split into 12 sets of 48, using a single processor to create 48 sets of input files and 48 model executables. These executables would then be run concurrently on 48 processors, each working on its own set of input data. When these runs were finished the data required for the analysis would be archived and all other input and output data deleted before starting the whole process again. The way that this is carried out is described in some detail below, in the spirit of the reader being able to repeat the experiments.

Before this, however, a brief description of the CMAQ directory structure will aid in understanding the process. The location of the CMAQ home directory is stored by the environment variable `M3HOME`. Under this the directories `M3HOME/data`, `M3HOME/models` and `M3HOME/lib` are referenced using the environment variables `M3DATA`, `M3MODELS` and `M3LIB` respectively. `M3DATA` contains sub-directories which hold all of the input and output data for the model runs, `M3MODELS` contains all of the CMAQ source code and `M3LIB` contains various libraries required for compilation of the CMAQ programs. A fourth directory, `M3HOME/scripts` contains all of the scripts used in building and running the model. A run script for a simple run to benchmark the model after compilation is supplied with the CMAQ source code and serves as a template for the user to modify for his/her own model runs. The `data` directory was duplicated when carrying out multiple runs and the `M3DATA` variable set to the location of a different `data` directory for each run. This is possible because of the way environment variables are inherited by sub-processes on a Linux system. If one thinks of a shell script as a process, then all programs and scripts called by that script are sub-processes of it. The processes and sub-processes have a structure similar to that of a directory tree, and values of environment variables are passed down that tree but not up it, so that an environment variable can have many different values at the same time on the system as a whole, but all processes downstream

of a point where it was given a particular value will only be aware of that single value. As well as the usual input and output files, a source code directory called `cb05cl_ae5_aq` was copied into each duplicate `data` directory. This contains source code files for the chemical mechanism and is normally found under `M3MODEL/include/release/`, but by making multiple copies of it in the `data` directories, different reaction rate and deposition velocity data could be included in the compilation of each CCTM executable.

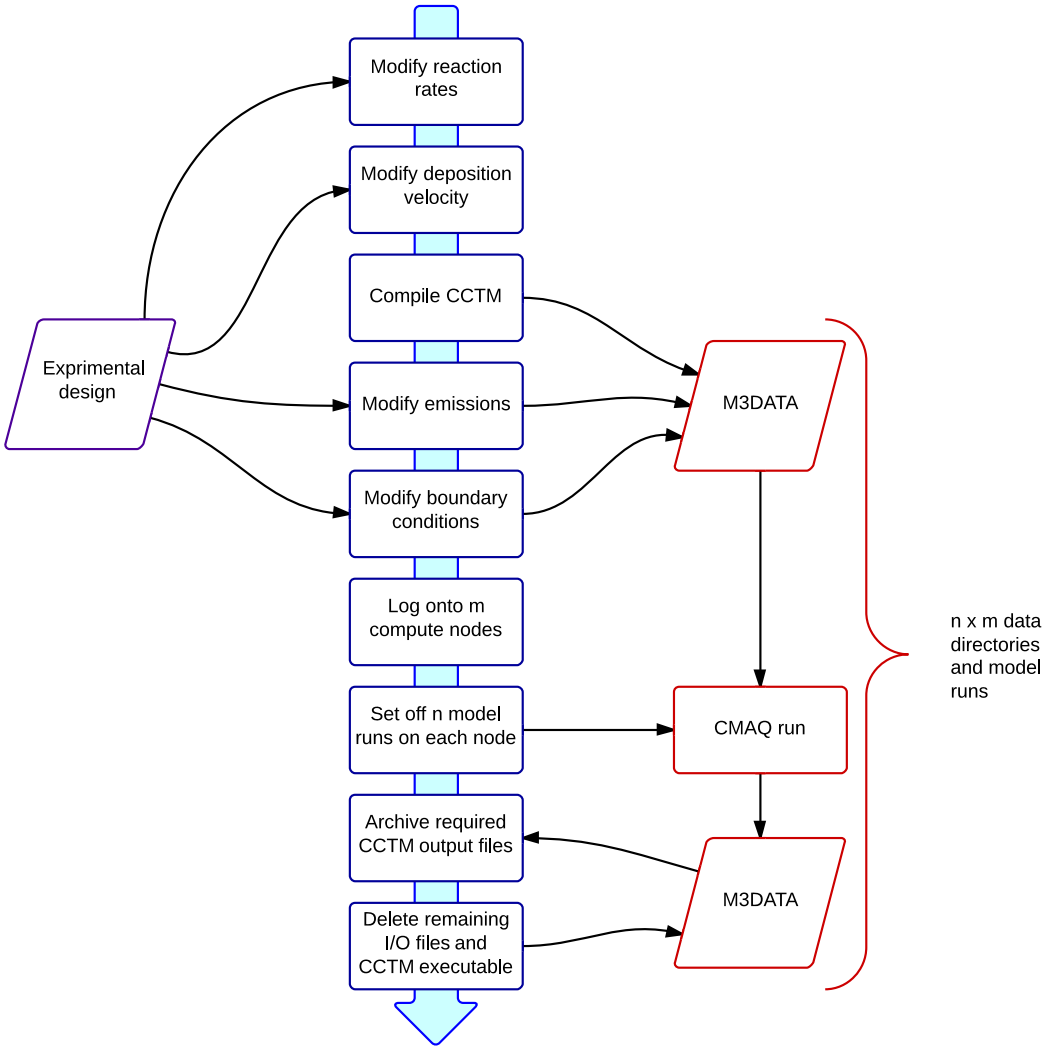


Figure 2.6: Schematic representation of the automated system to perform multiple CMAQ runs using n processors per node on each of m compute nodes

Figure 2.6 shows a conceptual representation of the system used to carry out large numbers of CMAQ runs according to a particular experimental design, which could either be a Latin hypercube or a Morris sampling plan. The processes

shown on the blue arrow are controlled by a master run script (appendix B.1.1) and proceed as follows:

1. Reaction rates are modified by changing the FORTRAN include file `RXDT.EXT` in each `M3DATA/cb05cl_ae5_aq` directory. This is achieved by concatenating a standard starting section of the file, a unique section containing modified Arrhenius coefficients created by the program `react3` (appendix B.1.4), and a standard ending section.

2. Ozone deposition velocities are modified by substituting a value from a pre-prepared text file into the file `GC_DEPV.EXT` in each `M3DATA/cb05cl_ae5_aq` directory.

3. Multiple CCTMs are compiled using a build script which has been modified to read include files from each `M3DATA/cb05cl_ae5_aq` directory instead of the central `M3MODEL/include/release/` directory. Each CCTM executable is then moved to the appropriate `data` directory.

4. All of the emissions files for a model run are copied into each `data` directory. The files are modified by running the R script `changeDomainEmis.R` (appendix B.1.2), which reads the values by which to modify each variable directly from the experimental design.

5. The boundary conditions used for the outer model domain are modified by running the R script `readProfile.R` (appendix B.1.3), which reads in the base case boundary conditions and values from the experimental design and produces a different boundary conditions file for each `data` directory.

6. Figure 2.6 refers to 'n x m' model runs, where m is the number of compute nodes to be used and n is the number of processors on each node. The master run script logs onto m compute nodes runs a script called run.seriesx, where x=1...m, (appendix B.1.5) on each one, which for each of the n processors sets the value of `M3DATA` to the location of one of the `data` directories

and launches the model run script `run.master10` (appendix B.1.6). This then launches the run scripts for `ICON`, `BCON` and the `CCTM` for each of the three model domains on all of the processors.

7. After waiting for the last output file from each of the model runs to appear in its appropriate directory, the master run script moves those output files which will be needed for the data analysis to an archive.

8. The remaining output files, the input files produced in steps 1,2,4 and 5, and all of the `CCTM` executables are deleted. The master run script then returns to step 1 and repeats the whole process again until the required number of model runs have been completed.

Chapter 3

Input factor Screening

This chapter describes the results of screening all of the gas-phase species in the emissions and boundary conditions (BCs) supplied to CMAQ, and all of the chemical reaction rates in the core CB-05 chemical mechanism, in order to identify which have the greatest influence over modelled concentrations of ozone and NO_2. As a variable which was considered essential to be included in any analysis of ozone in the CMAQ model, ozone deposition velocity was not included in the screening process. The results obtained while developing a method to adjust its value were deemed interesting enough to be presented on their own before moving on to the sensitivity analysis. First, however, initial conditions are considered with the intention of removing them from further analyses.

3.1 Initial Conditions

The effects of uncertainty in ICs are usually minimized when running CMAQ by using a 'spin-up' period, which is a period of modelled time at the start of the run from which no results are taken. Therefore, ICs are not considered in the main analyses described in this report. The recommended length of the spin-up period varies in the literature form two to ten days (Jiménez et al., 2007; CMAS, 2010), and the higher end of that range was chosen here. A test was carried out in which the ozone concentrations in the ICs were set at half and double their

baseline value, to make certain that this spin up period was long enough. The results of this are shown in figure 3.1, for a twenty day period, with the dashed vertical line indicating 240 hours of elapsed time. By this time the three lines on the graph have converged, and ten days seems to be an adequate spin-up. This is quite an extreme test, however, and the plot could be used to count back the number of days from the point where the lines converge, to give the appropriate spin-up for a set of ICs with a known confidence level.

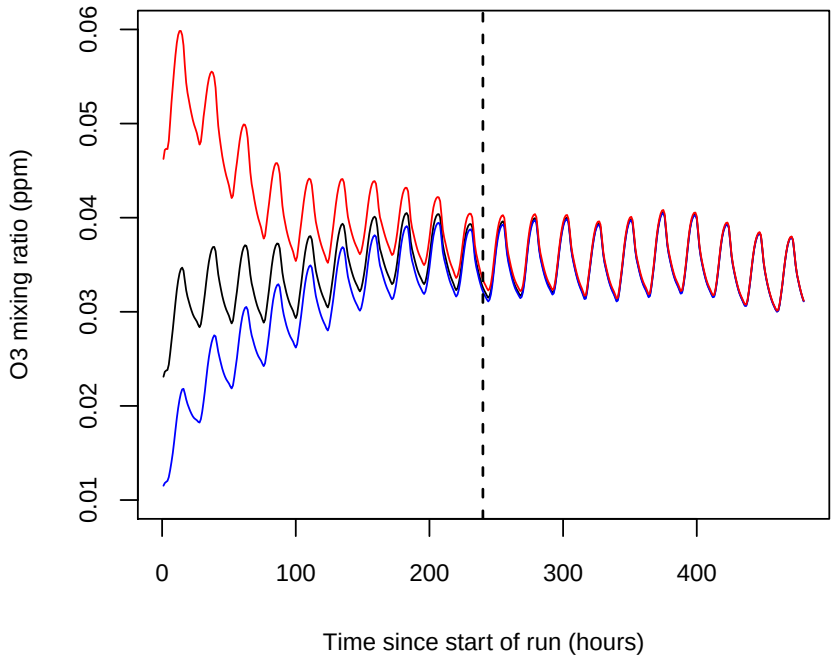


Figure 3.1: Modelled ozone concentration with ozone ICs at the baseline value (black), double this (red), and half of this (blue).

3.2 Screening with Morris' Method

This section describes the screening of emissions, BCs and reaction rates using Morris' method. The sensitivity analyses presented in chapter five also include ozone deposition velocity, but this was not included in the screening process as it was apparent from initial results that it was an important factor. It was expected that the model's sensitivity to various factors would not be the same across the whole domain, and although it was just the bottom layer of the inner domain

which was of primary interest, this still contains 9,360 grid cells. Performing the analysis for all of these cells would have been prohibitive so a subset was chosen. This subset comprised the 22 grid cells containing the air quality monitoring sites used in the DEFRA Phase 2 Regional Model Evaluation (Carslaw et al., 2013), the locations of which are shown in figure 3.2.

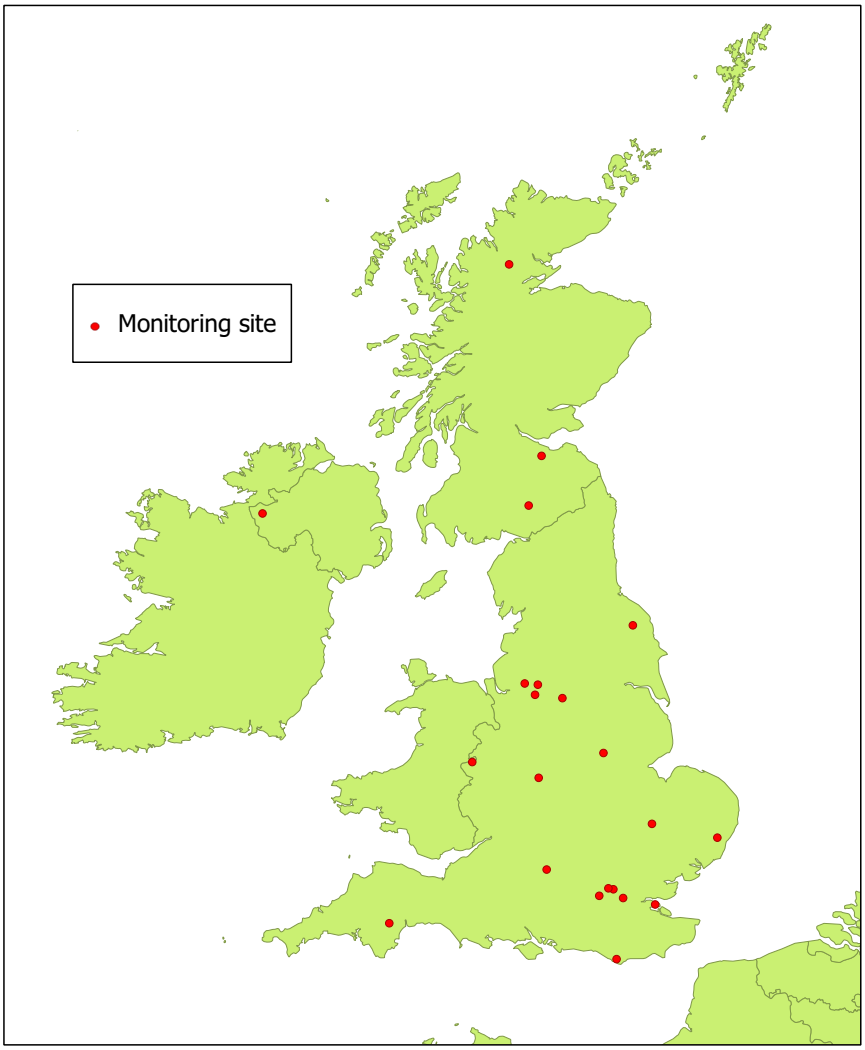


Figure 3.2: Locations of air quality monitoring sites which were used as locations for the screening process.

A total of 223 input variables were screened and ideally this would have been carried out in one operation. However, as the capability to perturb variables which are hard coded into the model source was developed as the project progressed, it was actually carried out in two stages, boundary conditions and emissions first, and then chemical reaction rates as a separate exercise. As a result,

the following two sections deal with each of these groups of variables separately.

3.2.1 Boundary Conditions and Emissions

Morris' method was used to screen 67 variables, 40 from the BCs and 27 from the emissions, for their influence on ozone and NO_x at 4 p.m. and midnight, on both the 10th May and 19th July. The BC variables are described in table 3.1, and have names which differ from those used in CB-05, as CMAQ 4.7.1 requires the user to supply BCs (and ICs) in the speciation used by the Second Generation Regional Acid Deposition Model (Stockwell et al., 1990).

Table 3.1: Boundary condition input variables screened with Morris' method

Variable name	**Description**
NO2	nitrogen dioxide
NO	nitric oxide
O3	ozone
HNO3	nitric acid
H2O2	hydrogen peroxide
HCHO	formaldehyde
PAN	peroxyacetyl nitrate
ETH	ethane
CO	carbon monoxide
ALD	acetaldehyde and higher aldehydes
ISO	isoprene
OLT	terminal olefins
OLI	internal olefins
TOL	Toluene and less reactive aromatics
XYL	Xylene and more reactive aromatics
SO2	sulphur dioxide
SULF	sulphuric acid
OP1	methyl hydrogen peroxide
OP2	higher organic peroxides
PAA	peroxy acetic acid
ORA1	formic acid
ORA2	acetic and higher acids
NH3	ammonia
HC3	alkanes, $2.7\times10^{-13} < k_{OH} < 3.4\times10^{-12}$

Table 3.1: (continued)

Variable name	Description
HC5	alkanes, $3.4\times10^{-12} < k_{OH} < 6.8\times10^{-12}$
HC8	alkanes, $k_{OH} > 6.8\times10^{-12}$
OL2	ethene
ACO3	acetylperoxy radical
TPAN	$H(CO)CH{=}CHCO_3NO_2$
HONO	nitrous acid
HNO4	peroxynitric acid
KET	ketones
GLY	glyoxal
MGLY	methyl glyoxal
DCB	unsaturated dicarbonyl
ONIT	organic nitrate
CSL	cresol and other hydroxy substituted aromatics
MACR	methacrolein
MVK	methyl vinyl ketone
BENZENE	benzene

The 27 gas-phase species from the emissions files, listed in table 3.2, are in the usual CB-05 speciation, so some represent individual species and some are lumped.

Table 3.2: Emissions input variables screened with Morris' method

Variable name	Description
ALD2	acetaldehyde
ALDX	propionaldehyde and higher aldehydes
BENZENE	benzene
CH4	methane
CO	carbon monoxide
ETH	ethene
ETHA	ethane
ETOH	ethanol
FORM	formaldehyde
HCL	hydrogen chloride
HONO	nitrous acid
IOLE	internal olefin carbon bond

Table 3.2: (continued)

Variable name	Description
ISOP	isoprene
MEOH	methanol
NASN	...
NH3	ammonia
NO	nitric oxide
NO2	nitrogen dioxide
SESQ	sesquiterpene
OLE	terminal olefin carbon bond
PAR	paraffin carbon bond
NR	small fraction of PAR
SO2	sulphur dioxide
SULF	sulphuric acid
TERP	terpene
TOL	Toluene and other monoalkyl aromatics
XYL	Xylene and other polyalkyl aromatics

Ten trajectories across the input variable space were used, meaning that 680 CMAQ runs were required for each day. At this stage uncertainties in input variables had not been assigned, so each variable was simply perturbed over a range of one half to double its baseline value. In both cases the subject day was the last day of a ten day simulated period. Figure 3.3 shows the results of one screening test for ozone, at Harwell, at 4 p.m. on the 19th July. μ^* is the mean of the absolute value of the elementary effects, and σ is their standard deviation. Some of the points have been labelled to show the most important factors at this particular time and place for this pollutant. However, if other such plots are made they are all different, meaning that as suspected the sensitivities do vary across time and location. With the screening being being done for both ozone and NO_2, for 2 times on 2 days at 22 locations, this would make a total of 188 plots, which is obviously too many to interpret. Instead, the highest μ^* value attained by an input factor from any screening was assigned to that factor, and these values used to discard those factors which fell below a threshold of 2 ppb

influence on either ozone or NO_2 output.

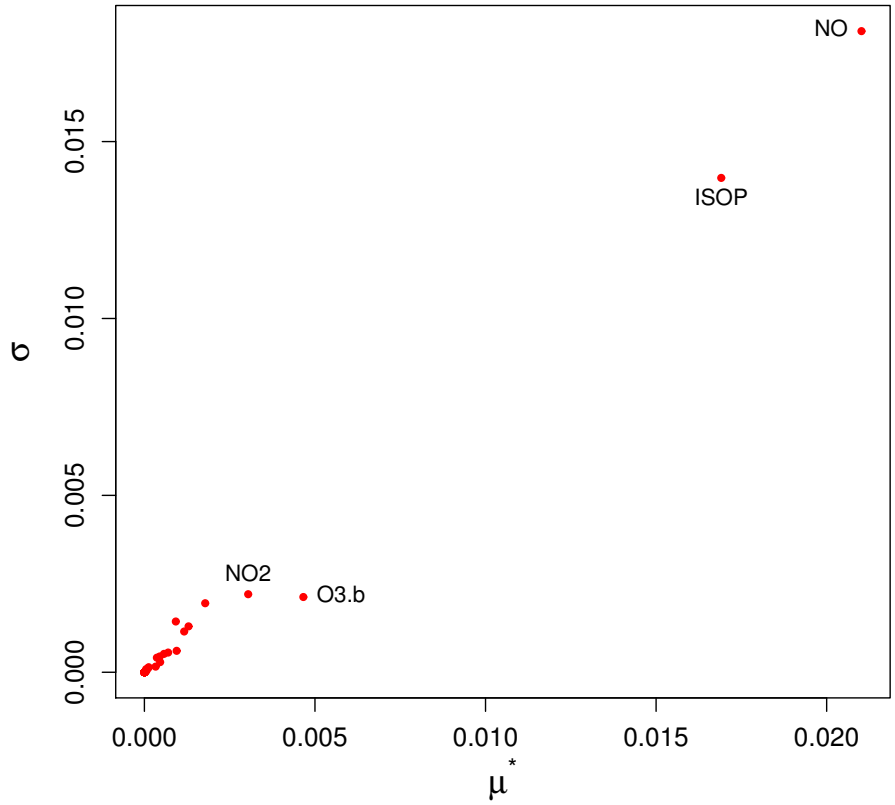


Figure 3.3: Results of Morris' screening method for ozone at Harwell, 4 p.m., 19th July. NO, NO2 and ISOP are NO, NO_2 and isoprene emissions. O3.b represents ozone boundary conditions.

3.2.2 Chemical Reaction Rates

The CB-05 core chemical mechanism contains 51 species and 156 reactions which are listed in Yarwood et al. (2005). CMAQ was actually compiled to run with the `cb05cl_ae5_aq` mechanism, which contains extensions for chlorine, aerosol and aqueous chemistry, but just the reaction rates in the core mechanism were considered to be relevant to modelled ozone and NO_2. Despite this, screening 156 factors for such a computationally expensive model is still an onerous task, and for this reason the number of trajectories used was reduced to five. Using a smaller number of trajectories will inevitably reduce confidence in the results, but

this number is still between the values of four and ten recommended by Saltelli et al. (2004). Nevertheless the range over which each rate was perturbed was reduced to ± 30 % of its baseline value. Five trajectories, each with 157 points, made a total of 785 model runs, which were carried out with a five day simulated time period. Most of the reaction rate calculations in the core mechanism are of the general form,

$$k = A \cdot e^{f(T)} \tag{3.1}$$

or

$$k = A \cdot g(T) \cdot e^{f(T)} \tag{3.2}$$

where $f(T)$ and $g(T)$ are constant at any given temperature, so that adjusting the coefficient A allows the rate to be perturbed while still maintaining its temperature dependence. A few of the reaction rate calculations have different forms, but all contain a constant which can be manipulated in order to change the rate while still maintaining the functional form of the equation. The photolysis rates can simply be multiplied by a factor before being modified to account for cloud cover within the CCTM.

Most of the reactions in CB05 have rates which increase with increasing temperature, and central England was around 10°C warmer on the afternoon of the 19th July than the 10th May. Also, 19th July had very low cloud levels and is nearer to the summer solstice than 10th May, so photolysis reactions would be at least as active. For these reasons, the screening was just carried out for 19th July, as the base case model run would have generally more active chemistry on this day.

In order to decide on a threshold level for μ^* at which to reject factors, some preliminary FAST experiments were carried out, in which NO and isoprene emis-

sions, and ozone boundary conditions were tested alongside the following reactions,

$$NO + O_3 \longrightarrow NO_2 + O_2$$

$$NO_2 + OH \longrightarrow HNO_3$$

which were chosen because of their importance in ozone production in previous studies, for example those by Beekmann and Derognat (2003) and Zhou et al. (2013). When compared in this way, these reactions were considerably less influential than the other factors, so the threshold for μ^* was set at 5 ppb with respect to ozone output. This subset of reactions also included all of the reactions which had μ^* greater than 2 ppb for NO_2, and is shown in table 3.3.

Table 3.3: Reactions retained after the screening process

CB05 label	**Reaction**
R1	$NO_2 + h\nu \longrightarrow NO + O$
R3	$O_3 + NO \longrightarrow NO_2$
R7	$NO_2 + O_3 \longrightarrow NO_3$
R9	$O_3 + h\nu \longrightarrow O_2 + O^1D$
R10	$O^1D \xrightarrow{M} O$
R11	$O^1D + H_2O \longrightarrow 2OH$
R28	$NO_2 + OH \longrightarrow HNO_3$
R30	$HO_2 + NO \longrightarrow OH + NO_2$
R66	$OH + CH_4 \longrightarrow CH_3O_2$
R74	$HCHO + h\nu \longrightarrow 2HO_2 + CO$
R87	$CH_3CO_3 + NO \longrightarrow CH_3O_2 + NO_2$
R112	$PAR + OH \longrightarrow 0.87XO_2 + 0.13XO_2N + 0.11HO_2 + 0.06CH_3CHO - 0.11PAR + 0.76ROR + 0.05ALDX$

The only expression here which requires more explanation is R112, where the complex form is due to the functional lumping employed in CB-05. This reaction represents the oxidation of paraffin bond hydrocarbons (alkanes) by the

hydroxyl radical, which in the carbon bond mechanism produces a complex mix of aldehydes and intermediate reaction products (Yarwood et al., 2005).

At the end of the screening process the reactions above were retained, along with CO and ozone BCs, and eight emissions variables. In the analyses which follow, the emissions inputs for the 9 km domain were perturbed separately to those for the 81 and 27 km domains, and ozone deposition velocity was added to make a total of thirty one variables. They are not listed in full here as this is done in chapter 4 when the distributions describing the uncertainty in each of the inputs are assigned.

3.3 Deposition Velocity

Dry deposition is the major physical mechanism by which ozone is lost from the atmosphere, and is modelled in the same way as meteorological quantities such as moisture and momentum in terms of the vertical flux in the lower layers of the atmosphere. This in itself involves several uncertain quantities, not least the parameterizations necessary to describe physical processes operating on spatial scales smaller than the vertical layer spacing of the model. Then additional uncertainties arise in calculating the resistance of various surfaces to ozone deposition. Following this the overall deposition velocity is calculated using an electrical resistance analogue, with different surface and aerodynamic resistances to deposition represented as resistors arranged both in series and parallel in an electric circuit (Pleim and Ran, 2011). As such, this part of the CMAQ model has various uncertain inputs which could be treated individually, but as this stage of the analysis, given time constraints and the level of complexity already involved, it was felt that it was better to treat ozone deposition velocity as a single input factor. The deposition velocity was adjusted as explained in the description of automated model runs in section 2.7, but in this initial experiment only eight different values were used, so the model source code was modified and compiled manually to produce eight different CCTM executables. Three locations are con-

sidered in this section; Harwell, in Oxfordshire, and London allow the comparison of the effect of deposition on ozone concentrations in rural and urban areas, and Strath Vaich in northern Scotland provides a good location to examine the effect of wind speed and direction on ozone deposition. These three locations are shown on the map in figure 3.4.

Figure 3.4: Locations of Harwell, London and Strath Vaich

3.3.1 Ozone Deposition at Harwell and London

Time series of ozone deposition and concentration are shown for Harwell in figure 3.5 and London in figure 3.6. In both figures ozone concentration is on the top graph and deposition, in kg ha^{-1}, on the bottom, for various values of deposition velocity expressed as a percentage of the baseline value. Blue lines show lower deposition velocity, and red lines higher, hence the colours are reversed in the top and bottom graphs as higher rates of ozone deposition lead to lower atmospheric concentrations and vice versa. This is exactly the behaviour that would be expected, demonstrating that the method for adjusting the deposition velocity was

working correctly.

Examining the vertical scales of the bottom graph in both figures reveals that ozone deposition was considerably greater, in terms of absolute value, at Harwell than in London. Urban ground surfaces are much less amenable to ozone deposition than rural ones, with stomatal uptake by vegetation being a major ozone sink (Andersson and Engardt, 2010). This has a corresponding effect on ozone concentrations in the top graph of each figure, the vertical spacing of the red and blue lines being much greater at Harwell than in London. As a consequence, uncertainties in deposition velocity will have a much greater effect on atmospheric ozone concentration in rural areas than urban ones.

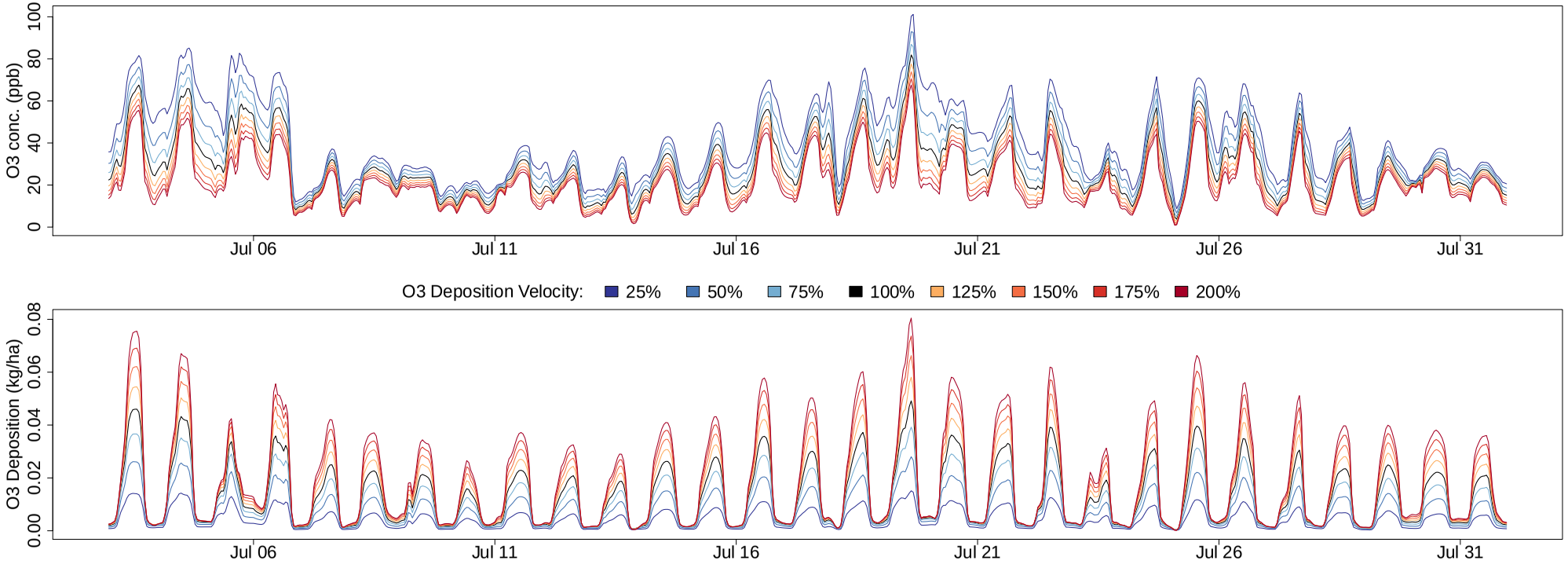


Figure 3.5: Time series of ozone concentration (top) and ozone deposition (bottom) for different deposition velocities, Harwell, 3rd-31st July.

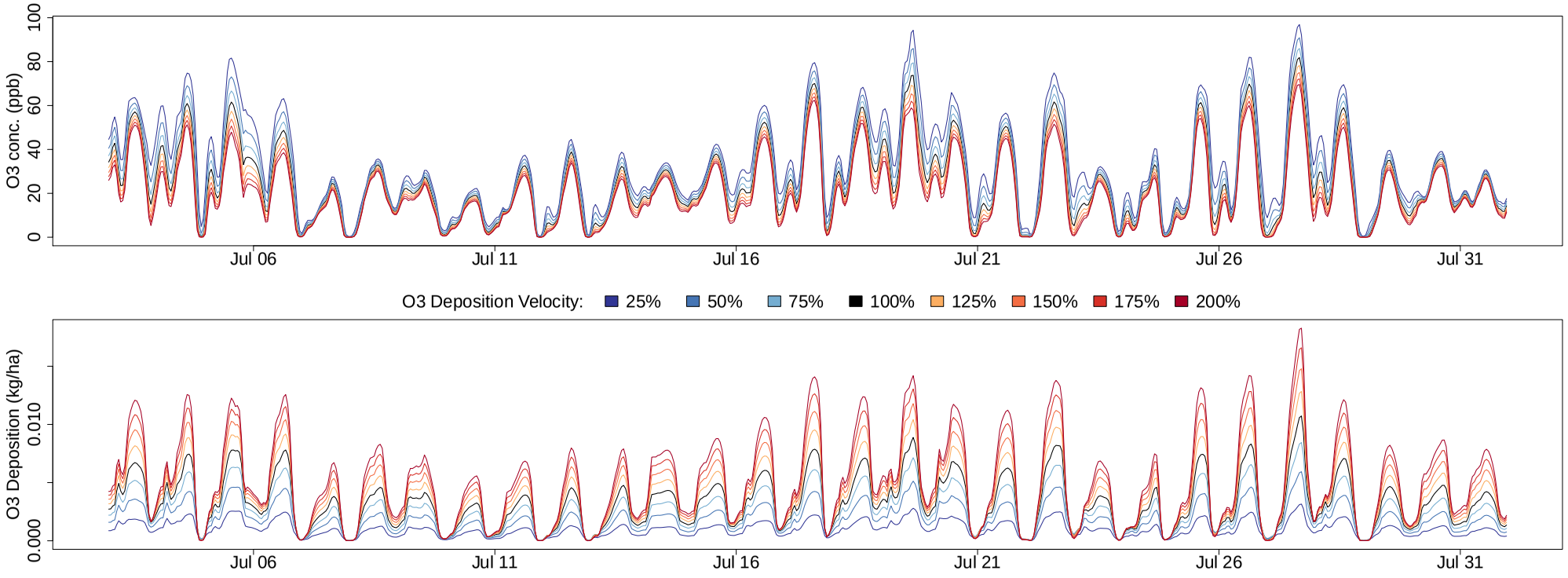


Figure 3.6: Time series of ozone concentration (top) and ozone deposition (bottom) for different deposition velocities, London, 3rd-31st July.

3.3.2 The Effect of Wind Speed and Direction on Ozone Deposition

Being far from any major urban centres, Strath Vaich provides a good location to examine the effect of wind speed and direction, without the confounding effects of differing land use. Inspection of the the top graph in figure 3.7 reveals that there was a period during the first half of the month when the atmospheric concentration of ozone was largely unaffected by deposition velocity. Despite this, the absolute values of ozone deposition, in the bottom graph, are clearly affected by the same changes in deposition velocity. This may at first seem hard to reconcile, until one considers that the top graph shows the ozone concentration in air, which is moving through the grid cell which contains Strath Vaich, and the bottom graph shows the deposition to the ground in that grid cell, which is stationary. Wind speeds were higher during this period than later in the month, and when the air is moving more quickly it spends less time in each grid cell, and less time in total depositing ozone in all grid cells it has passed though before reaching its current location. This means that deposition velocity is less important in determining atmospheric concentrations so that changing its value becomes ineffectual.

The effect can be seen more explicitly in figure 3.8, where hourly wind speeds and directions are plotted as wind roses conditioned on the change in ozone concentration caused by deposition. For each hourly model time step the vertical separation of the top blue line and the bottom red line on the top graph of figure 3.7 indicates the effect that changing the deposition velocity from 25 % to 200 % of the baseline value has on atmospheric concentrations. Those values are then split by quartiles, and the corresponding wind speeds and directions plotted on separate wind roses. The top left wind rose shows the lowest quarter of hours, in terms of difference between ozone concentrations at the highest and lowest values of deposition velocity, and the bottom right wind rose the highest quarter. Wind speeds gradually become lower as the effect of deposition velocity

on ozone concentrations becomes higher. Also apparent is the change in wind direction across the plots, with the highest deposition effects being coupled with more southerly winds. A quick glance back at the map in figure 3.4 shows that air coming from this direction travels further over land than air travelling from the west, and therefore the opportunity for ozone deposition to vegetation is increased.

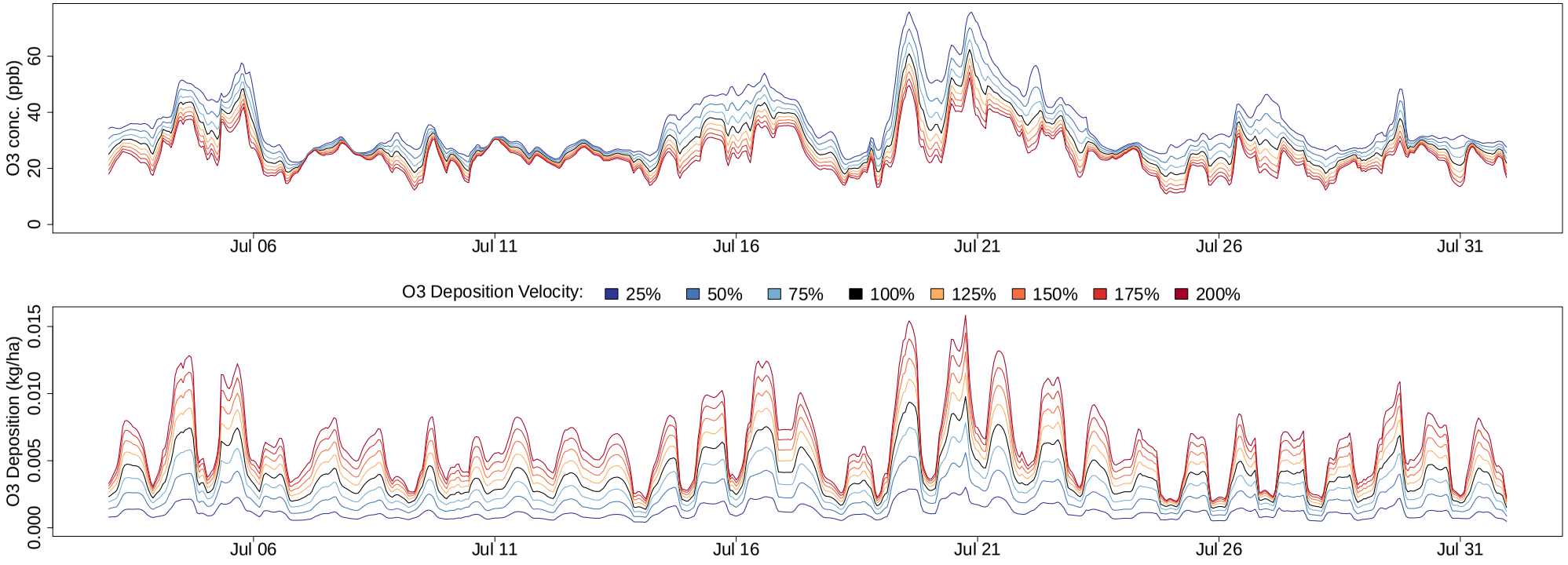


Figure 3.7: Time series of ozone concentration (top) and ozone deposition (bottom) for different deposition velocities, Strath Vaich, 3rd-31st July.

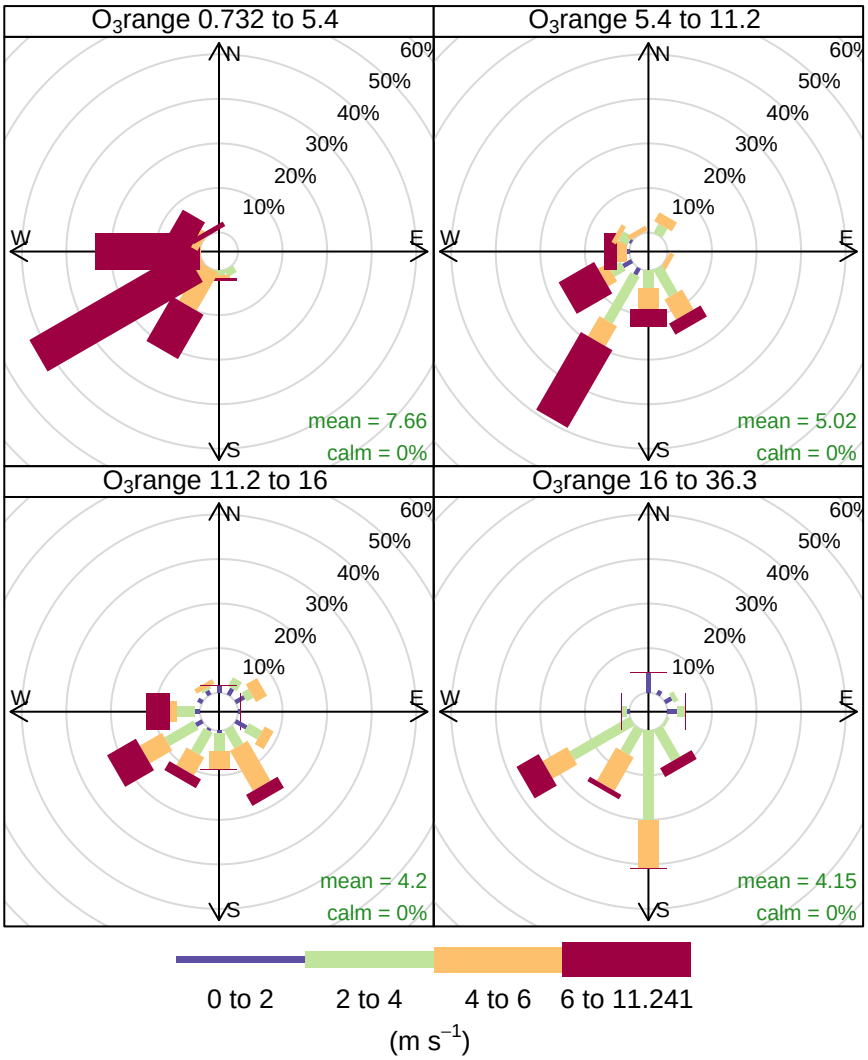


Figure 3.8: Wind rose plots for Strath Vaich, 3rd-31st July, conditioned on ozone deposition.

3.4 Uncertainties in Meteorology

Meteorological conditions generally are an important factor in determining levels of atmospheric pollution, so it follows that errors in the meteorological data which are fed into an air quality model may cause errors in the pollutant concentrations which are output by that model. Meteorological variables have not been included in the analyses presented in this report, however, for several reasons. The first is the practical aspect of respecting mass conservation in modelled chemical species

when perturbing variables concerned with atmospheric dynamics such as wind speeds. Recalling the description of CMAQ in chapter 1, the core of a Eulerian CTM is a set of continuity equations which describe the movement of mass into and out of each grid cell and its neighbouring cells, and as wind is moving mass, randomly perturbing its speed would upset the balance between these equations. This is a problem which Hanna et al. (2001) assert can be minimized by varying wind speed uniformly across the whole domain, but it would seem that even if this does alleviate the mass conservation issue other problems would still remain. There are many other variables output by a meteorological model such as WRF and ingested by CTMs which are deterministically linked to wind speed, such as Monin-Obukhov length and planetary boundary layer height, so simply changing the wind speed without also modifying these variables appropriately will lead to modelling the air quality in a physically unrealistic scenario. Similar arguments can be made for other key meteorological variables such as temperature and humidity, changes in which affect other physical quantities in the real world which would not be reflected in the data fed into the CTM.

The WRF model output used in the analyses described throughout this report had been augmented using a data assimilation technique known as 'grid analysis nudging'. The data used for this was NOAA Global Forecast System (GFS) historical archived data (NOAA, 2015) and the assimilation nudges the WRF output towards this data by adding terms to the model equations in a process described as 'Newtonian relaxation' (Hahmann et al., 2010). Importantly, this occurs as the model is running, so avoids the inconsistencies with other variables mentioned above. There is no pretence that this will provide meteorological inputs to CMAQ which are a perfect representation of reality, but it is fair to assume that most of the time large scale synoptic conditions will be correct, even though small scale errors may still exist.

Meteorological data assimilation is a highly active research area and efforts towards its improvement are being constantly made by the WRF model developers

in the US, who are generally well funded. Also, historical meteorology reanalysis data sets are increasingly being made freely available at higher spatial resolution. It is hard to imagine more resources being directed towards improving any of the other sources of uncertain input data used by CMAQ, so in the context of a sensitivity analysis being used to direct future research efforts, it is less important to include Meteorology than emissions, for example.

Chapter 4

Sensitivity Analysis

This chapter describes sensitivity analyses of CMAQ ozone and NO_2 output concentrations to uncertainties in those input variables which were retained following the screening process described in chapter 3. The first step in this process was to successfully emulate the model, so this is described in section 4.1. In section 4.2 the estimated distributions of the uncertainty in the input factors are given and the results of the FASTs carried out over these distributions are displayed. Section 4.3 complements the FASTs with plots which show the sensitivity of model output to changes in the value of one particular input, whilst accounting for the uncertainty in other inputs. Section 4.4 illustrates the sensitivity of ozone concentrations to changes in NO_x and VOC emissions with the use of bivariate contour plots.

4.1 Emulation of the CMAQ Model

As stated in previous chapters, rigorous sensitivity and uncertainty analysis techniques cannot be performed directly with the CMAQ model because long run times make the large number of runs required by such techniques infeasible. Whilst emulation was carried out using standard techniques and tools, as described in chapter 2, the emulators used here are amongst the most ambitious described in the literature in terms of numbers of input variables, and represent

the first implementation of Gaussian process emulation of a complex Eulerian air quality model.

4.1.1 Performing the Training Runs

A 31 dimensional Latin hypercube sampling plan was created with uniform distributions for each dimension bound by zero and one. This was then scaled to be bounded by one half and two, to provide factors by which to multiply the input variables so that each would be sampled from half to double its baseline value. This range was larger than the ranges of the input uncertainties to be used in the sensitivity analyses, so provided scope to widen them, or to perform other experiments as required.

For those inputs which take the form of spatio-temporally varying fields, namely emissions and deposition velocity, the whole field was multiplied by the same factor. This serves both to maintain mass balance within the model, and also to keep the whole analysis tractable. Also, as described in the screening section, the reaction rate scaling factors where applied in such a way as to allow spatio-temporal variation due to temperature fluctuations across the domain. The model output to be emulated was either ozone or NO_2 concentration in a single grid cell at a single time step. Where it was required to emulate the output at multiple locations or time steps then separate emulators were built for each. To be clear, this means that scalar valued model output for a single point in space and time was emulated with respect to changes in inputs across the whole spatial domain, and through time from the start of the run until the output time.

Throughout the project the aim was to construct emulators which were accurate enough to allow their mean predictions to be used with confidence in place of CMAQ. The emulator variance is used, however, in history matching, but this process is also most effective when the emulator variance is as small as possible. To facilitate the construction of accurate emulators a nominal target of 20 training runs per variable was chosen, with the exact number being dictated by

the efficient use of the number of available processors. The cluster used had 96 processors on its compute nodes, and at times all of these could be used, or sometimes just half, depending on other users needs, so the total number of runs was always divisible by 48 or 96. All of the results presented use emulators with 31 input variables, so Latin hypercubes with 576 points were used, and hence 576 training runs performed, equating to 18.6 runs per variable.

Because of the high volume of data involved in making such large numbers of runs, a maximum run length of eleven simulated days was used. Sometimes consecutive simulated time periods were used, such as when running the whole of July 2006, which involved three sets of 576 runs, two for 10 days and one for eleven days. In these circumstances it was vital to ensure that the numbering of archived output data directories and the numbering of points in the Latin hypercube remained consistent throughout.

The R package *DiceKriging* was used to build all of the emulators used in this project, as described in section 2.2, because of its speed and accuracy.

4.1.2 Example output

Figure 4.1 shows the CMAQ model output of ozone concentrations throughout the 21 day period used for many of the analyses presented here, along with the output of emulators giving the response to changes in 31 input variables as described above. CMAQ produces output at hourly time steps, so for the 21 days, 504 emulators were used for each location. The Latin hypercube design does not include the base case model run, so this is not included as training data used to build the emulators. Therefore, reproducing the base case CMAQ output is not an easier task for the emulators than reproducing the output given by any other combination of model input values, at least within the same region of the input variable space.

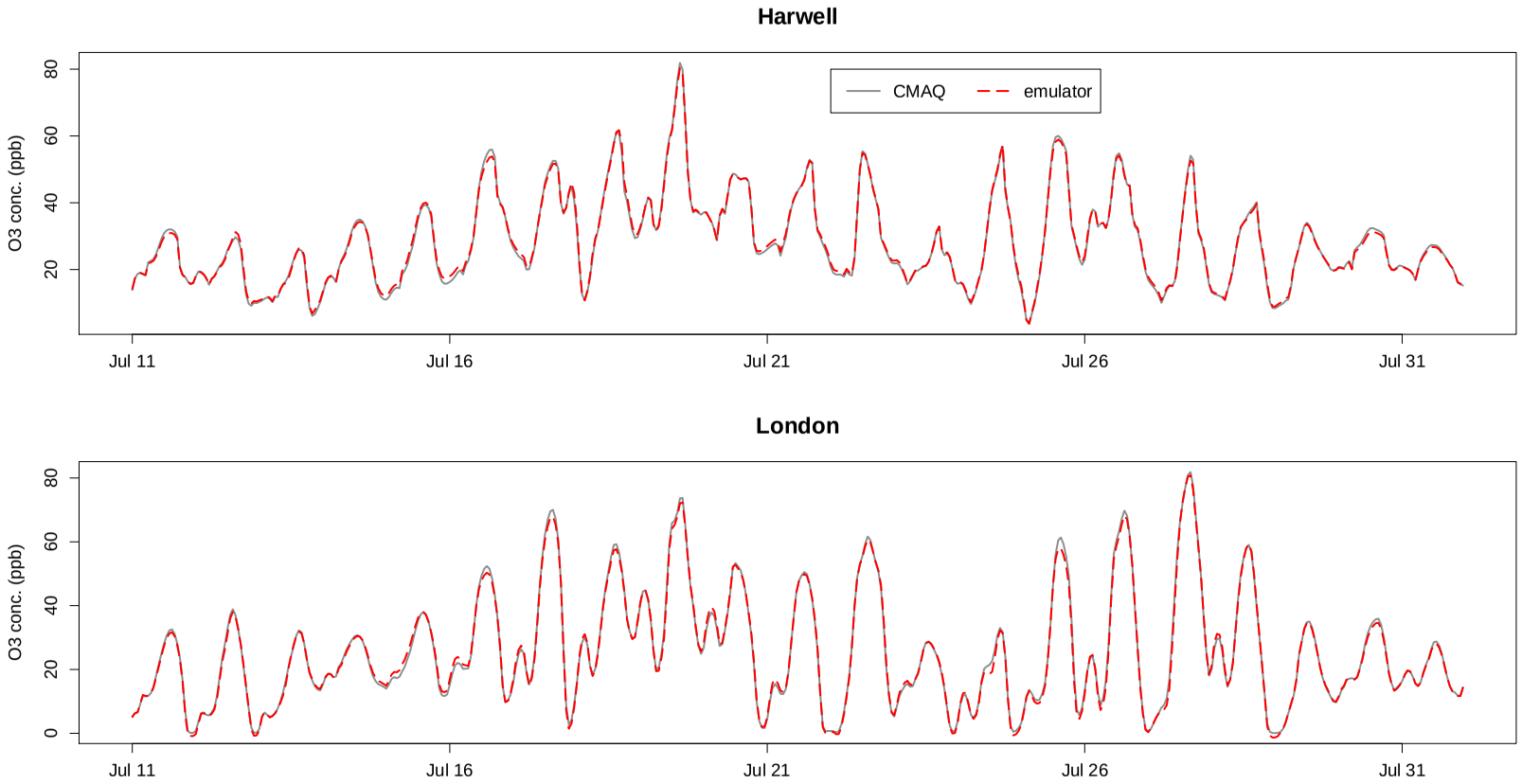


Figure 4.1: CMAQ and emulator ozone concentrations at Harwell and London, 11th - 31st July 2006.

4.1.3 Validating the Emulators

Figure 4.2 shows a standard set of emulator diagnostic plots which are produced by the *DiceKriging* package. The plots shown are for an emulator of CMAQ ozone output at 4 pm on the 19th July at Harwell, a particular time and location which is referred to frequently throughout this report. The 'leave one out' analysis is performed by building the emulator without one of the training runs and then using the left out run as validation data with which to compare the emulator output at that same set of inputs.

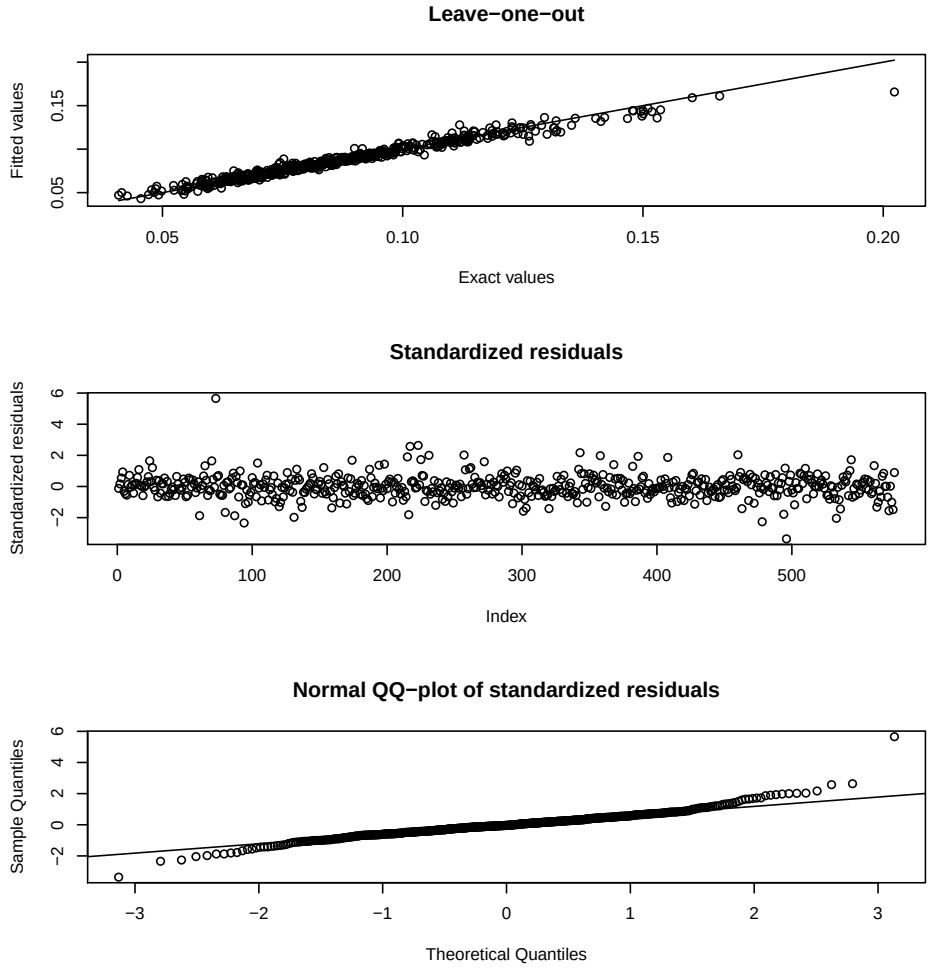


Figure 4.2: Standard emulator diagnostic plots produced by the *DiceKriging* package.

The top plot shows the emulated values of the CMAQ model output plotted against the actual output values so that a perfect emulator would produce

points which all lie on the 1:1 line. The fit appears to be reasonably good, with just one outlier corresponding to a CMAQ output value of around 200 ppb (the axes on the plot are in ppm), which is so far from the observed value of 97 ppb and the base case modelled value of 79 ppb, that it is likely to be output produced at a physically unrealistic point in the input variable space. The middle plot of figure 4.2 shows the errors between the model output and the emulator output as standardised residuals, that is, divided by the emulator standard deviation. Bastos and O'Hagan (2009) recommend that most of these points should lie within a range of $\pm$ 2, but a small number outside of this range can be ignored, so given that there are 576 points in total, the plot would seem to show acceptable performance, especially as Bastos and O'Hagan also state that too small a spread around zero would indicate that the emulator variance has been overestimated. The bottom plot in figure 4.2 shows a normal quantile - quantile plot of the same standardised residuals, which do deviate from the 1:1 line, but only near its ends, and in any case, O'Hagan (2006) states that deviation from the normality assumption which is inherent in the use of Gaussian processes is of little consequence provided enough training runs are made.

Bastos and O'Hagan also suggest plotting these standardised errors against the emulator outputs and against the inputs as further graphical diagnostics. The standardized errors are shown plotted against the emulator outputs in the left-hand plot of figure 4.3, and the unstandardised errors are shown in the right-hand plot because it is useful to the modeller to see the errors in physically interpretable units. The vertical red line in both plots shows the base case emulator output, and for around 20 ppb either side of this the spread of the errors is fairly constant. Moving further away from the base case, however, the spread becomes greater, particularly above 120 ppb. This heteroscedasticity in the errors indicates non-stationarity (Bastos and O'Hagan, 2009), which violates one of the assumptions underlying the Kriging process, but given that the output concentrations in this region of the plot are well above both the base case model output and the observed

value this might not be a problem.

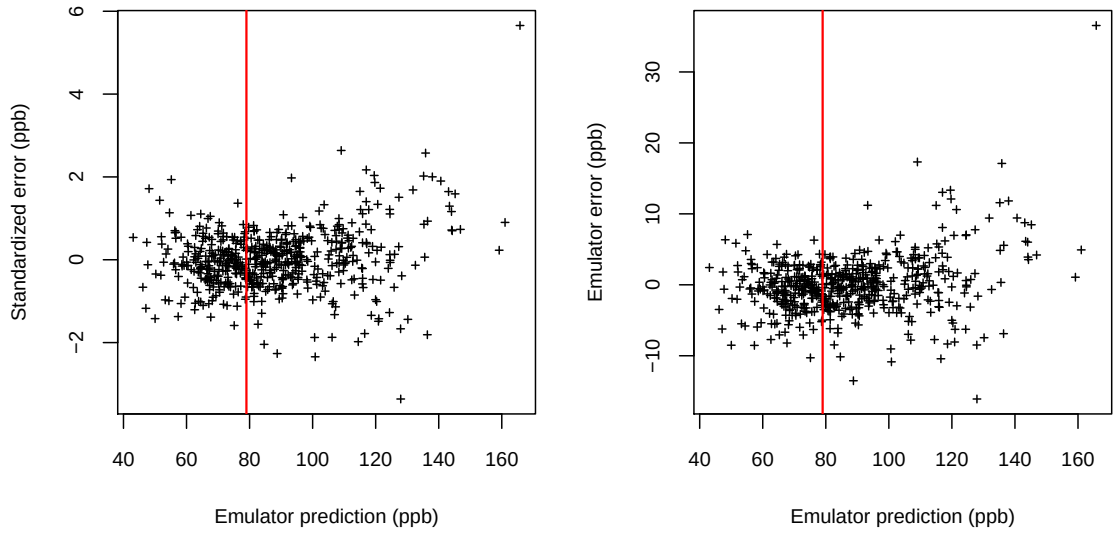


Figure 4.3: Standardised (left) and unstandardised (right) emulator errors plotted against emulator predictions for leave one out cross validation. The red lines show the emulated base case run.

To investigate this further the errors are plotted against each of the inputs in figure 4.4. The standardised and unstandardised plots in figure 4.3 have a different pattern as the emulator standard deviation used to standardize each error has a different value. The difference between the left and right-hand plots is not great, however, and the most extreme outliers are present in both cases. For this reason, just the unstandardised errors are plotted in figure 4.4, which for all of the inputs lie in a horizontal band as Bastos and O'Hagan state would be expected for a correctly functioning emulator. The small number of outliers are again apparent but there is nothing systematic about the errors so this is considered to be satisfactory.

The validation presented so far is for an emulator of CMAQ ozone output at a single point in space and time, but the results presented hereafter use emulators of both NO_2 and ozone, for each hourly time step in a 21 day period, and for various locations. In order to allow the appraisal of a greater number of emulators it is useful to summarize the validation of each one using the the mean absolute error (MAE) of the emulated output, which in the case above was 2.80 ppb.

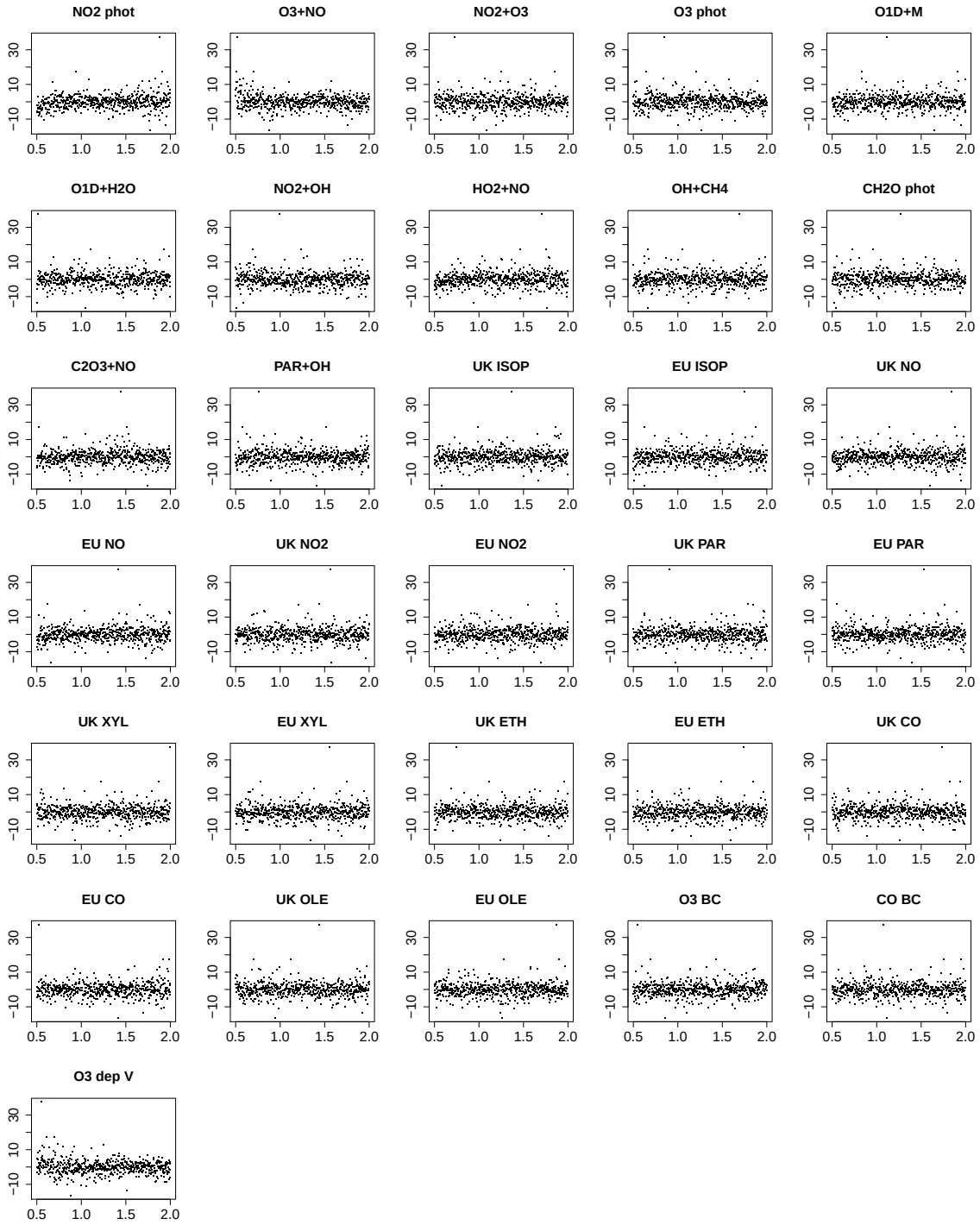


Figure 4.4: Emulator prediction errors, in ppb on the vertical axes, plotted against the factor used to scale each model input.

The leave one out cross validation procedure, as implemented in the *DiceKriging* package, does not re-estimate the emulator parameters each time a different training run is left out. In a companion package, called *DiceEval*, a K-fold cross validation procedure is implemented, which when K is set to the same number as the number of training runs, equates to a leave one out cross validation (Dupuy and Helbert, 2013). In *DiceEval* the parameters are estimated every time a new emulator is made with a different training run left out, and consequently the validation is more rigorous, but takes much longer to perform, which has implications for the number of emulators which can be validated. In order to resolve this issue, the analysis above was performed with both packages, and the emulated ozone concentrations given by the emulators produced by each package with the same training run left out were plotted against each other in figure 4.5.

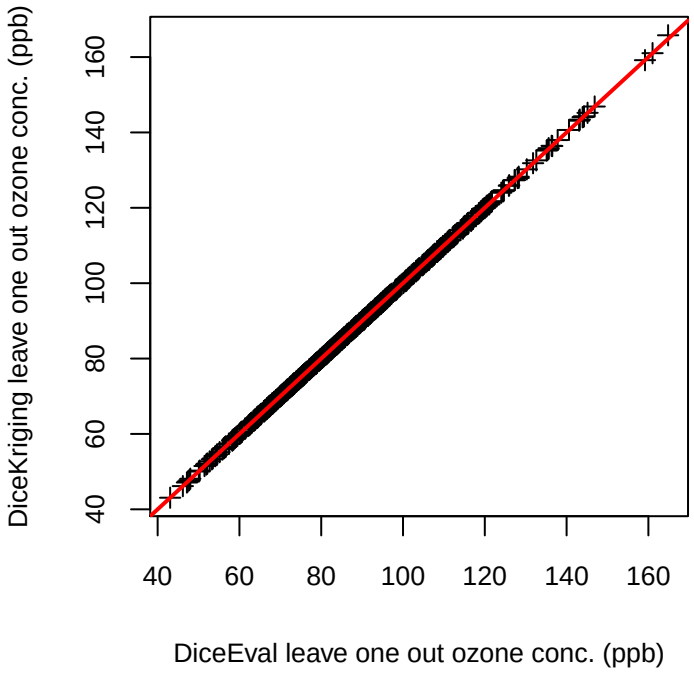


Figure 4.5: Ozone concentrations from emulators produced during different leave one out cross validation procedures in the *DiceKriging* and *DicEval* packages.

There is an almost exact match in the emulator output from each package and the MAEs calculated from the results of each differed by only 0.02 ppb, indicating

that it was safe to proceed in validating a larger number of emulators with the leave one out cross validation procedure from the *DiceKriging* package. Roustant et al. (2012) suggest that the dangers of not re-estimating the emulator parameters are most severe when only a small number of training runs are available, so it would appear that it has not been an issue in this case because of the large number of CMAQ runs which were performed.

The MAEs for emulators of ozone and NO_2 concentrations were calculated for 125 randomly chosen time steps and locations, and their distributions are shown by the box plots in figure 4.6. The median MAE for ozone emulators is around 1.3 ppb and three quarters have MAEs below 1.9 ppb, while for NO_2 over three quarters of the the emulators have a MAE of below 0.5 ppb. The MAE for the emulator which was subject to more detailed analysis above is shown by the red circle, and it is clear that the majority of ozone emulators and all of the NO_2 emulators perform better than this. These results were considered adequate to proceed with using the emulators for the analyses which follow.

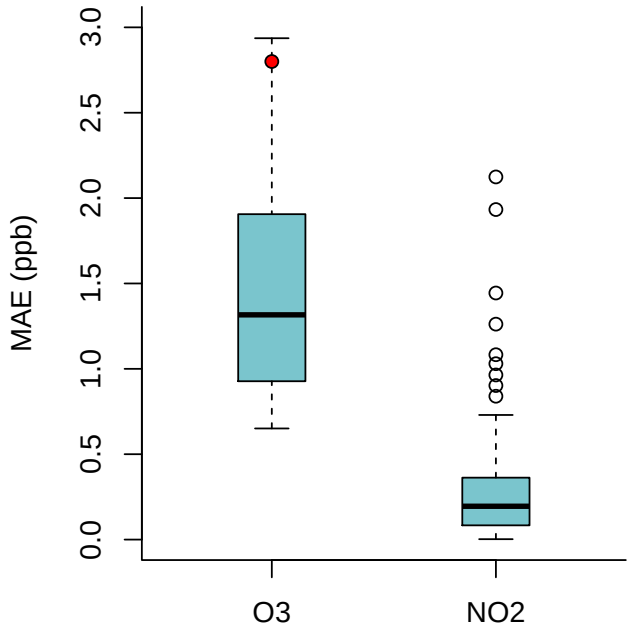


Figure 4.6: Distributions of the mean absolute error (MAE) statistic for leave one out cross validation of emulators for 125 randomly chosen times and locations.

4.2 Fourrier Amplitude Sensitivity Tests

The uncertainty distributions for the model inputs are given in table 4.1. Normal distributions have a mean of 1.0, and uniform distributions are centred on 1.0, so that values sampled from them equate to a factor by which to multiply the appropriate input variable. The reaction rate uncertainties are taken from the NASA/Jet Propulsion Laboratory Atmospheric Chemical Data Evaluation (Sander et al., 2011), and emissions uncertainties from the NAEI uncertainty estimation given by Passant (2003). This estimation was created using a spread sheet add-on package for Monte Carlo analysis to propagate uncertainties in emissions factors, activity rates and various other data through to national emissions totals. Passant gives uncertainty estimates resulting from this numerical technique of ± 7 % in NO_x emissions and $+11$ % / -9 % in VOC emissions, the same technique applied to the 2006 inventory gave an estimate of ± 10 % for both NO_x and VOCs. The same document gives a much larger estimate of ± 30 % for NO_x and VOCs derived from expert opinion. This larger figure has been included in the present work, as there are major assumptions used in the Monte Carlo analysis, such as the conjecture that official statistics are 'subject to very limited uncertainty', and that missing sources of emissions would not contribute significantly to the total. The results of the FASTs show emissions as either 'UK' or 'EU', with UK representing perturbations in the 9 km domain, and EU representing perturbations in the 27 and 81 km domains together. In the absence of available information, the same level of uncertainty was assigned to BCs and to ozone deposition velocity as to emissions.

Table 4.1: Distributions used to characterise input uncertainties (all centred on 1.0)

Variable description	**Distribution**
R1: NO2 photolysis	normal, sd=0.2
R3: O3 + NO	normal, sd=0.1
R7: NO2 + O3	normal, sd=0.15
R9: O3 photolysis	normal, sd=0.3

Table 4.1: (continued)

Variable description	Distribution
R10: O1D + M	normal, sd=0.5
R11: O1D + H2O	normal, sd=0.08
R28: NO2 + OH	normal, sd=0.4
R30: HO2 + NO	normal, sd=0.3
R66: OH + CH4	normal, sd=0.1
R74: HCHO photolysis	normal, sd=0.4
R87: C2O3 + NO	normal, sd=0.5
R112: PAR + OH	normal, sd=0.3
NO emissions	uniform, ± 0.3
NO2 emissions	uniform, ± 0.3
ISOP emissions	uniform, ± 0.3
PAR emissions	uniform, ± 0.3
XYL emissions	uniform, ± 0.3
ETH emissions	uniform, ± 0.3
CO emissions	uniform, ± 0.3
OLE emissions	uniform, ± 0.3
O3 BCs	uniform, ± 0.3
CO BCs	uniform, ± 0.3
O3 deposition velocity	uniform, ± 0.3

Most of the plots in this section show time series of total effects for specific locations for the 21 day period from 11th to 31st July inclusive. These were made by constructing a separate emulator of either CMAQ ozone or NO_2 output at that location for each of the 504 hourly time steps in that time period. A separate FAST was then carrried out with each emulator before combining the results to produce the plots. This is a time consuming process, and so was parallelised by using 21 processors to each work on a 24 hour section of the data before collating the results and producing the plot. For each plot, data was read from 12,096 separate netCDF files in order to produce the 504 emulators, and the following 504 FASTs required over three million emulator 'runs' - a task obviously impossible with CMAQ itself. UK wide average plots were made by emulating the mean pollutant concentration over land in the 9 km model domain, and as

such include a small amount of Eire and a small corner of north east France. It should also be remembered that plots for a specific location are actually for the whole 9 km grid square which contains that location.

Showing the cumulative total effects as a function of time was first suggested by Saltelli et al. (1998), who normalise the indices so that their total sums to one and hence the numbers on the vertical axis represent a true fraction of the total variance. Here the indices were not normalised, which means that the total on the 'fraction of variance' scale often adds up to more than 1.0, and this can be seen as an indicator of the total amount of interaction between the input factors. To clarify, the effect of an interaction between two factors, for example, is included in the total effect of both factors, and similarly the effect of an interaction between three factors would be included in the total effect of all three. The FASTs were carried out for all 31 variables identified in the screening process, but only those which accounted for more than 1 % of the variance at any single time step are shown on the time series plots. This means that the legend for each plot contains some different input factors, although the most important ones are usually the same. Each plot has between 16 and 18 variables, and due to the difficulty in generating contrasting colour schemes with this many shades, some of the colours will vary from one plot to the next, although again the main factors are usually the same colour. In any case, the vertical order of the variables on the plot corresponds to the ordering of its legend from top left to bottom right. It is reassuring that all of the time series plots only show a subset of the 31 factors, as this shows that the threshold was not set too high in the screening process.

The results should not be over interpreted by asserting that a high total effect for a particular input infers that the input is a real world driver of levels of the pollutant. The primary aim is to apportion model output uncertainty into relative contributions from different sources, thereby identifying improvements which could be made to reduce overall output uncertainty. This relies on the assumption that the input uncertainties are independent, even when the absolute

values of those inputs may sometimes not be so. Having said this, however, a modelling practitioner with a high degree of confidence in the model, and a good working knowledge of the physical and chemical processes being modelled, may choose to assign causal relationships between high model sensitivities and high pollution levels, particularly if that reinforces prior beliefs about the nature of the pollution episode.

4.2.1 Ozone

Figure 4.7 shows the time series of total sensitivity indices for modelled UK average ozone concentration. Of particular note is the dominance of BCs at the start and end of the period, when wind speeds were relatively high. Deposition velocity becomes more important as wind speeds drop in the middle of the period, as suggested in the previous chapter. Also, as the weather becomes warmer in this middle period chemical processes start to dominate - BCs are important when there is little photochemical activity and vice versa. The diurnal variation in sensitivity to NO_2 photolysis, shown in yellow at the bottom of the plot, can be clearly seen. It is important to note that this does not go to zero, even at night, as although the value of the total effect gives the sensitivity at just that particular time step, this sensitivity is to variations in the rate for the whole period. To clarify this point further, at 2 am, for example, NO_2 photolysis has ceased - the rate is zero, so when it is perturbed by multiplying it by a scaling factor, it is still zero. However, the ozone concentration at 2 am is affected by the concentration during the previous day, and as this is sensitive to the NO_2 photolysis rate and hence so is the concentration at 2 am.

Interestingly, the model output is not particularly sensitive to uncertainty in emissions in this UK wide plot, but the same is not necessarily true at specific locations, as can be seen on the plot for Harwell in figure 4.8. Here the model is much more sensitive to NO emissions, particularly those from the UK. Comparing the two plots generally, the UK wide plot is much smoother in its temporal

variation than the Harwell plot, but the overall patterns are fairly similar. This is probably due to Harwell's location as a rural background monitoring site in central England. One such pattern is the quenching of excited oxygen atoms (O1D + M) shown in red and ozone photolysis immediately below in purple becoming important factors around the middle of the time period. A glance back at the time series of ozone concentrations at Harwell in figure 2.5 shows that this was a period of slightly lower concentrations in between two pollution episodes. Staying with figure 2.5 for a moment, it can be seen that the model has the greatest difficulty in reproducing observed concentrations during the afternoon peaks, particularly when concentrations are highest. Figure 4.8 shows that uncertainty in reaction rates taken as a whole dominate the sensitivity analysis at these times.

As a test of the sensitivity of the method to the assigned input uncertainty distributions, which are after all only estimates, the FASTs for Harwell were repeated with all variables given uniform distributions spanning ± 10 % of the baseline value, and the results are shown in figure 4.9. The effects of emissions, BCs and deposition velocity remain the same with respect to each other, which is unsurprising as their uncertainty distributions have all changed by the same amount. The relative importance of these inputs taken as a whole, compared to the importance of reaction rates taken as a whole, also does not change significantly. What is affected, however is the relative importance of different reaction rates to each other, as they have varying levels of uncertainty in the original analysis, so their distributions have all changed by different amounts. These facts do demonstrate the importance of attempting to correctly specify the input uncertainties, but the test was quite severe, with some uncertainty ranges changing by a factor of three or more, so smaller mis-specifications may not be as important.

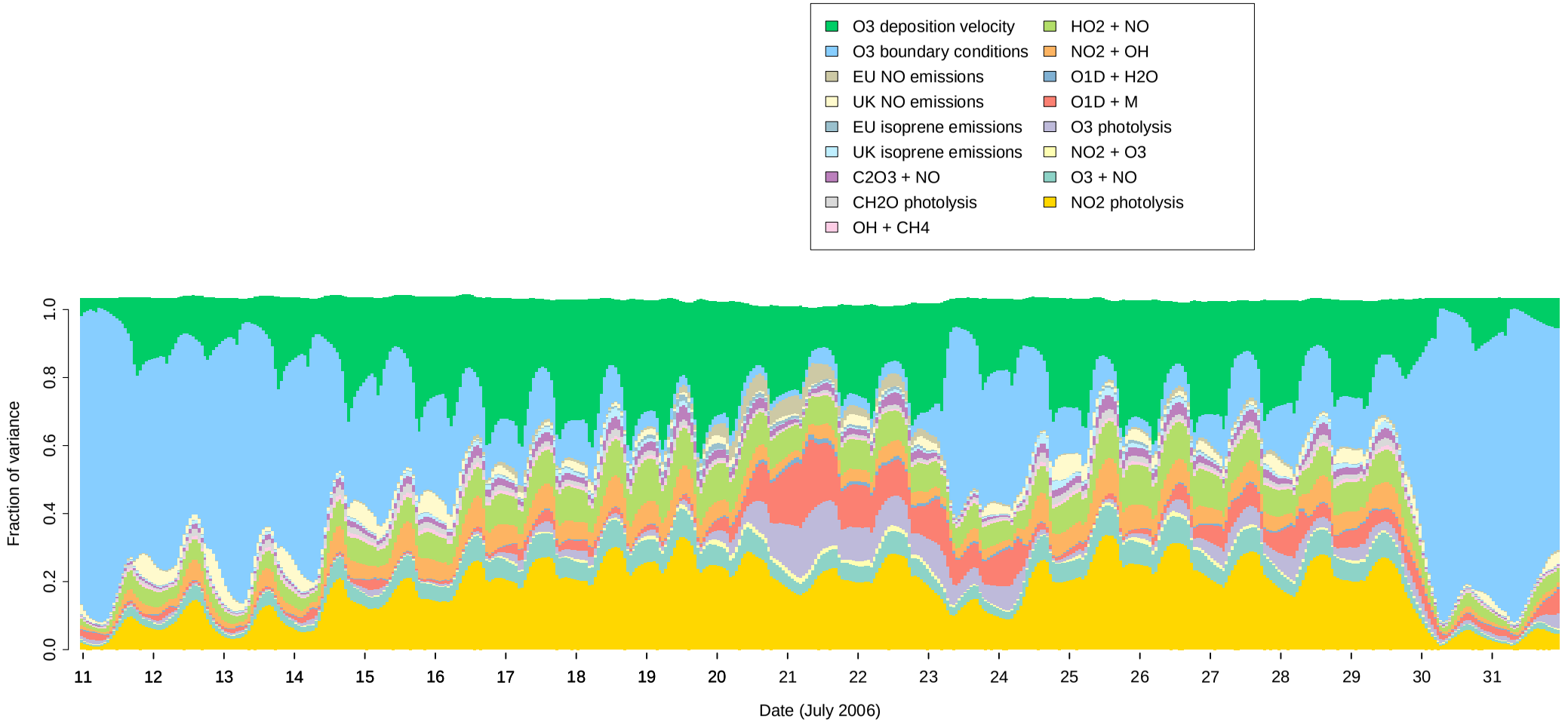


Figure 4.7: Time series of FAST total effect indices on modelled UK average ozone concentrations, 11th-31st July

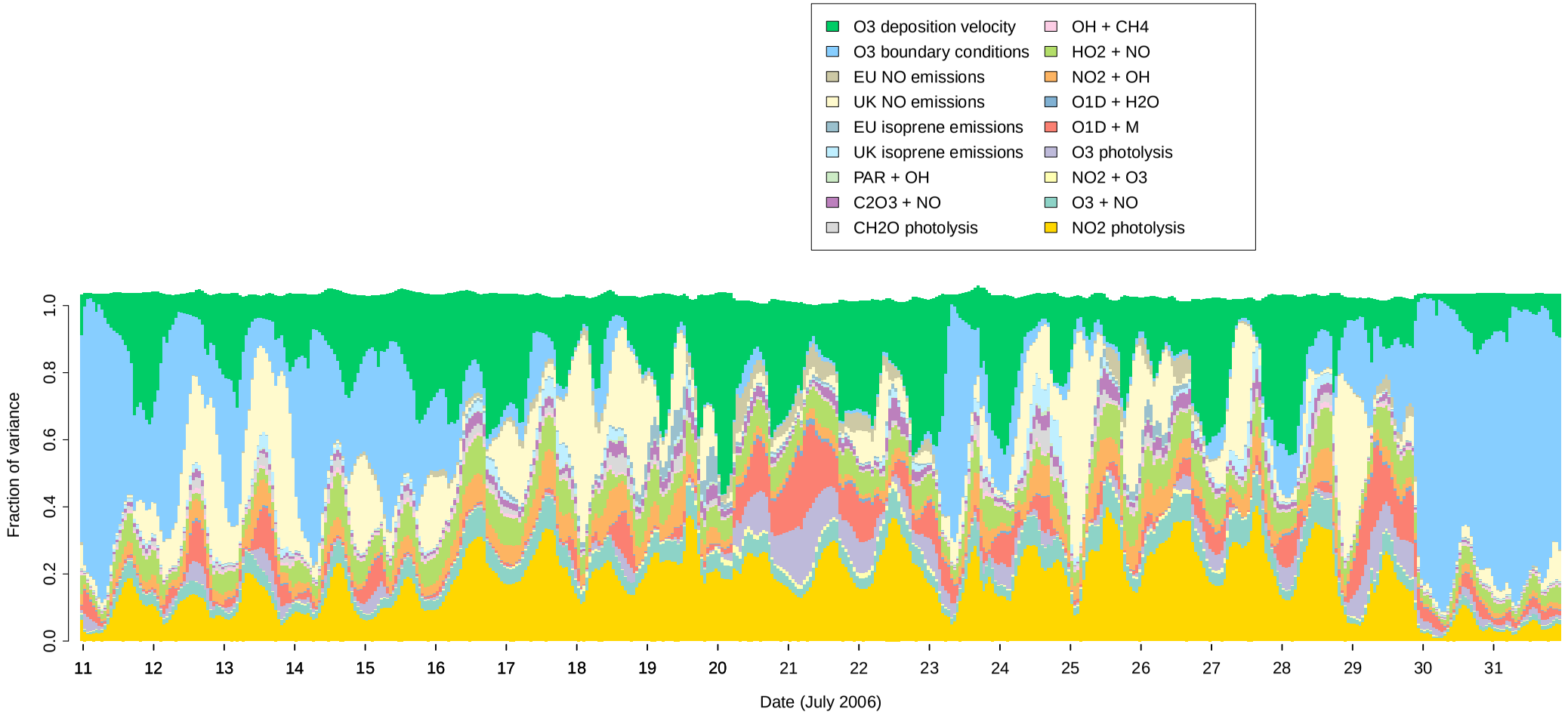


Figure 4.8: Time series of total effects on modelled ozone concentrations at Harwell, 11th-31st July

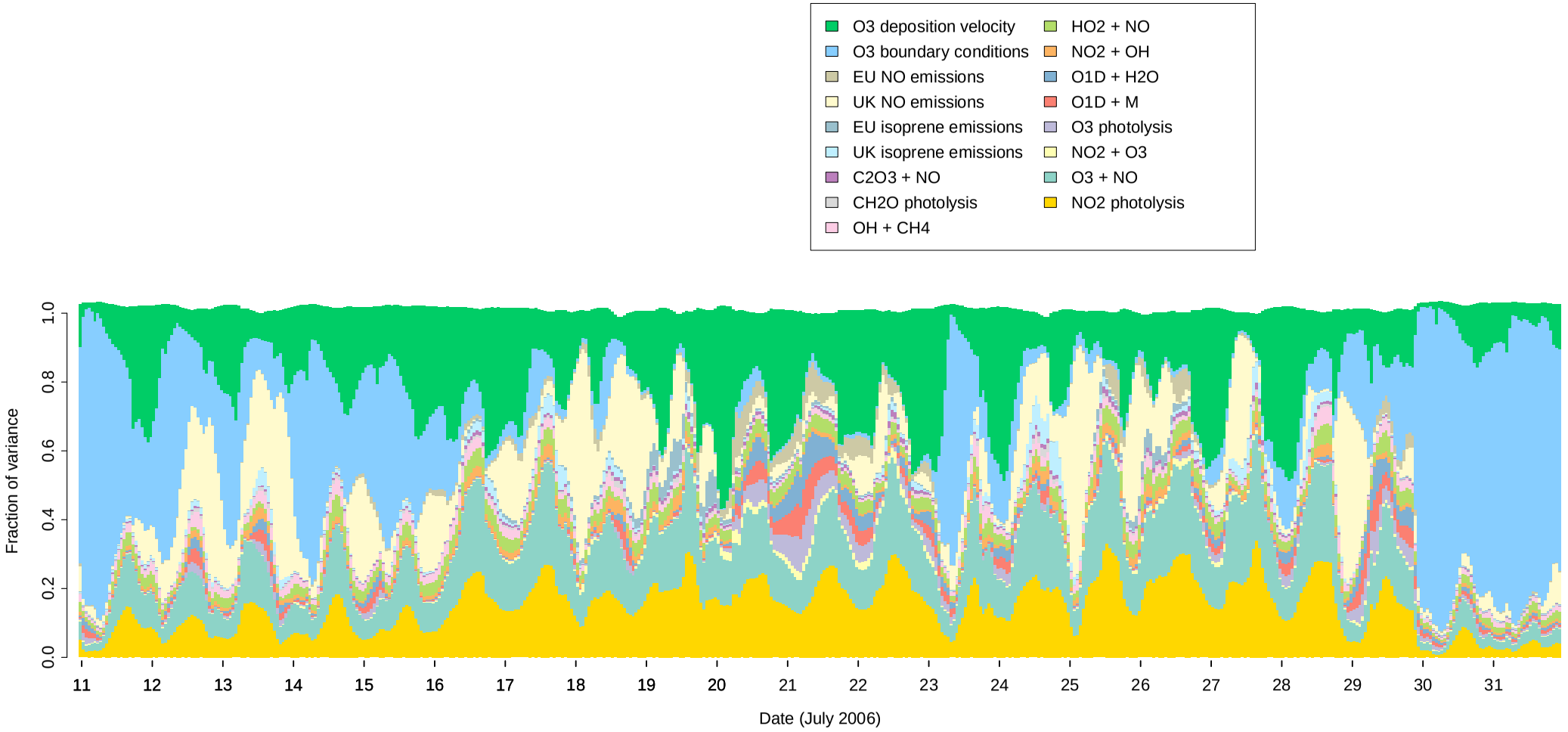


Figure 4.9: Time series of total effects on modelled ozone concentrations at Harwell, 11th-31st July, with uniform input distributions bounded by $\pm 10\%$ of the basecase model run.

Figure 4.10 shows the results for a grid square in London which contains the South Kensington and Bloomsbury air quality monitoring sites. These sites are classed as 'urban background', and it is useful to think of the analysis in this way - the grid cell size means that any urban area in this domain would be modelled as such given that the spatial resolution would not capture such features as roadside effects. In comparison to the previous plots, NO emissions are highly important, and are likely to be dominated by local traffic emissions. What is also apparent is that the diurnal peaks in the influence of reaction rates are even more marked, and in combination with the high traffic emissions and heatwave weather conditions this is suggestive of the model simulating a Los Angeles style photochemical smog.

In all of the analyses so far the only VOC which has featured on the plots (and hence the only one accounting for more than 1 % of the variance at any time) is isoprene, a biogenic species emitted by vegetation, particularly in hot weather conditions. Isoprene emissions are widespread accross most land areas and so it is not surprising that they are more influential than anthropogenic VOCs in the UK wide and Harwell plots, as emissions of anthropogenic species tend to be much more localised. London has the greatest anthropogenic emissions in the 9 km domain, however, so it is a surprise their uncertainties do not feature in the analysis.

A further feature worthy of note in figure 4.10 is the more 'spikey' nature of the top surface of the plot, indicating that interactions between variables become important for short, isolated periods of time. Closer inspection reveals that these points are shortly before midnight in every case, and given that this would also be shortly after sunset at this time of year, these interactions could be related to the transition from the daytime photochemical regime to night time chemistry. One of these periods is examined in more detail in figure 4.11, where the overall height of the bars represents the total effect as the sum of the main effect and interactions. Uncertainty in UK NO emissions can be seen to be dominant, and

the high degree of interaction with the other factors is clearly evident.

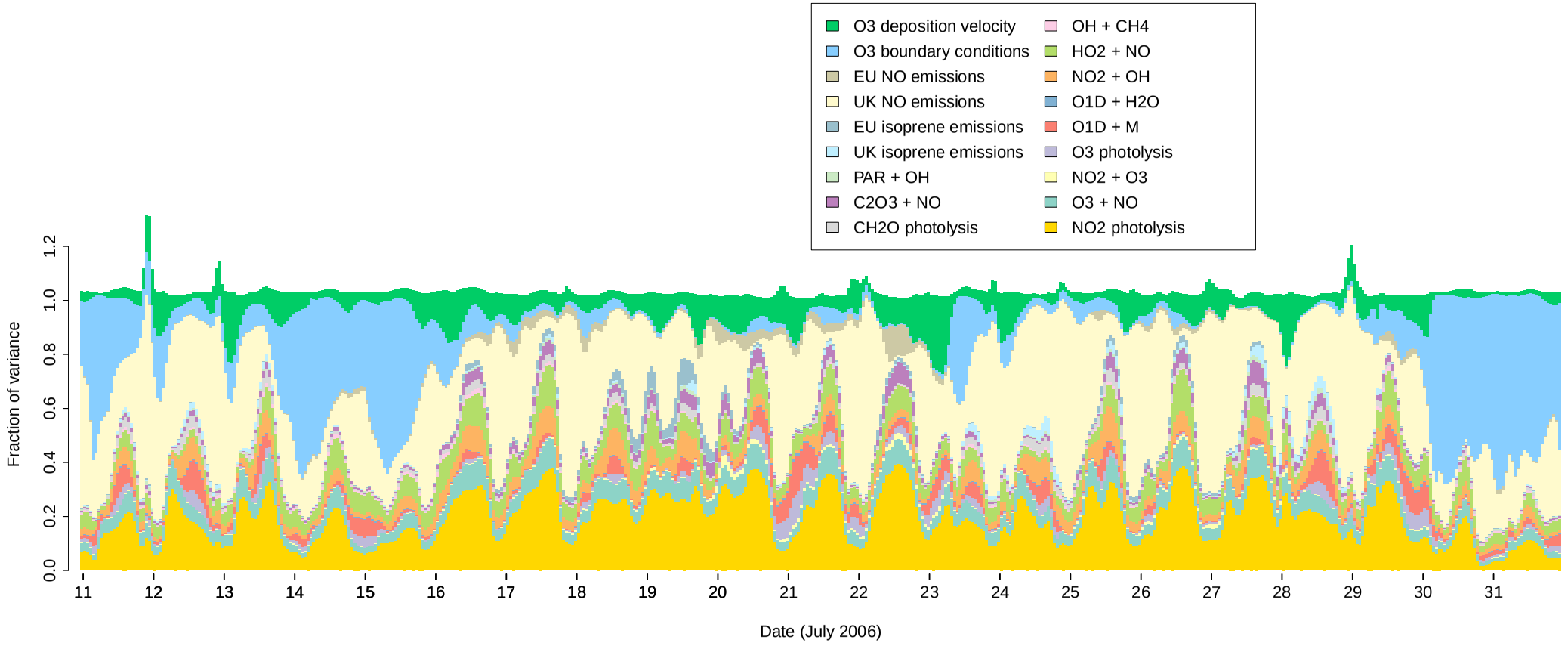


Figure 4.10: Time series of total effects on modelled ozone concentrations in London, 11th-31st July

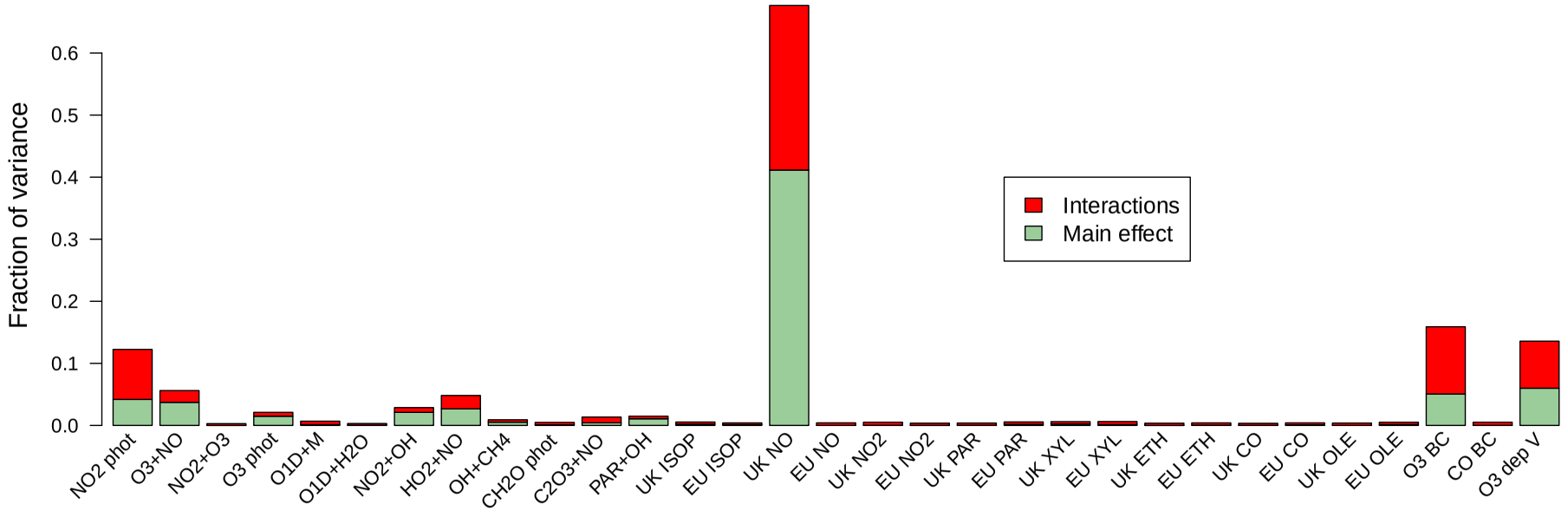


Figure 4.11: Main effects and interactions for ozone sensitivity in London, 10pm on 11th July

4.2.2 Comparison of July 19th and May 10th

A comparison of the sensitivity of peak afternoon ozone on May 10th and July 19th to all 31 input factors is shown in figure 4.12. These two days were chosen because they had qualitatively similar weather conditions, but July 19th was 12 K warmer in the afternoon than May 10th. The ozone peak occurred at 4 pm on the former day and 3 pm on the latter. May 10th was 42 days before the summer solstice, so would have received a similar, but slightly smaller, amount of solar irradiance to 19th July, which was 28 days after the solstice. The days leading up to 10th May had around 20 - 50 % cloud cover over the south of England, while July was largely cloudless for the whole month. This would affect the amount of solar radiation reaching the lowest parts of the troposphere and also accounts for the 12 K difference in temperature shown in table 4.2. The variables in table 4.2 are averaged over the six hour period running up to the ozone peak, apart from wind direction, which is given as a range. The two days had similar synoptic circulation patterns, with moderate Easterly or South-Easterly winds bringing air masses from continental Europe.

Table 4.2: Meteorological variables at Harwell for the six hours preceding peak ozone concentrations on the 10th May and 19th July

variable	**10th May**	**19th July**
wind speed (m s^{-1})	4.11	5.41
wind direction (°)	60 - 91	115 - 154
temperature (K)	293	305
water vapour mixing ratio (kg kg^{-1})	0.00916	0.00955

The greater amounts of solar irradiance on the 19th July are reflected in the greater importance of uncertainty in NO_2 photolysis seen in figure 4.12. The higher temperature generally makes uncertainty in chemical reactions, in the bottom portion of the graph, more important on 19th July than 10th May. Also worth noting, and also due to temperature, is the increased significance of isoprene emissions in July. Isoprene has strong ozone production potential (Derwent et al.,

1996), and its emissions from vegetation have a positive exponential relationship with air temperature (Laffineur et al., 2011). Ozone boundary conditions can be seen to play a much larger role on 10th May, and earlier in this chapter, when discussing the 21 day time series of sensitivities, such an effect was said to be due to higher wind speeds. However, deposition velocity is also more important on 10th May despite the fact that section 3.3.2 gave an illustration of such effects being due to lower wind speeds. In any case the wind speeds were similar on 10th May and 19th July and the higher importance of these two factors together on the 10th May is likely to be due to the lower temperatures and reduced chemical activity as compared to the 19th July. The reduced importance of modelled chemistry means that perturbing the reaction rates has a smaller effect on the output, causing other factors take up more of the variance decomposition. This demonstrates the value of performing a global sensitivity analysis over perturbing one or just a few factors at a time. Another possible explanation for the greater importance of deposition velocity on 10th May is that vegetation may have had more open stomata due to greater photosynthetic and growth activity and lower heat and moisture-deficit stress than in July (Emberson et al., 2013).

Figure 4.12: FAST for peak afternoon ozone at Harwell on 10th May and 19th July.

4.2.3 Comparison of Market Harborough and London

A similar comparison to the one in the previous section is presented in figure 4.13, but this time for two locations, Market Harborough, in Leicestershire, and London, both for the afternoon ozone peak on 19th July. The purpose of this is to compare the model sensitivity at a time of elevated ozone levels at an urban and rural location, as the grid cell used is that which contains the Market Harborough air quality monitoring station, which is outside of the town itself.

The first point to note is the greater sensitivity to deposition velocity at the rural location, which is likely to be caused by a greater absolute value of ozone deposition, due in turn to a higher proportion of vegetated land surface to which ozone can readily deposit than in London. The only emissions sensitivities of any significance are for NO and isoprene, both for the EU and the UK. Interestingly, the sensitivity to isoprene emissions uncertainty is geater in London than Market Harborough, even though the absolute amount of these emissions is smaller in the general area in and around London. This could be because isoprene is so potent in terms of ozone production potential that perturbations in its emissions when the absolute amount is small can have a great effect on model output by reducing its concentration to ineffectual levels. Alternatively, it could be related to the fact that both locations are more sensitive to uncertainty in EU than UK isoprene emissions, and London is closer to mainland Europe. Even though the e-folding lifetime of isoprene with respect to breakdown by reaction with ambient concentrations of OH is only about 30 minutes (Jacobson, 2002), it is the breakdown products which go on to contribute to ozone production further down the line.

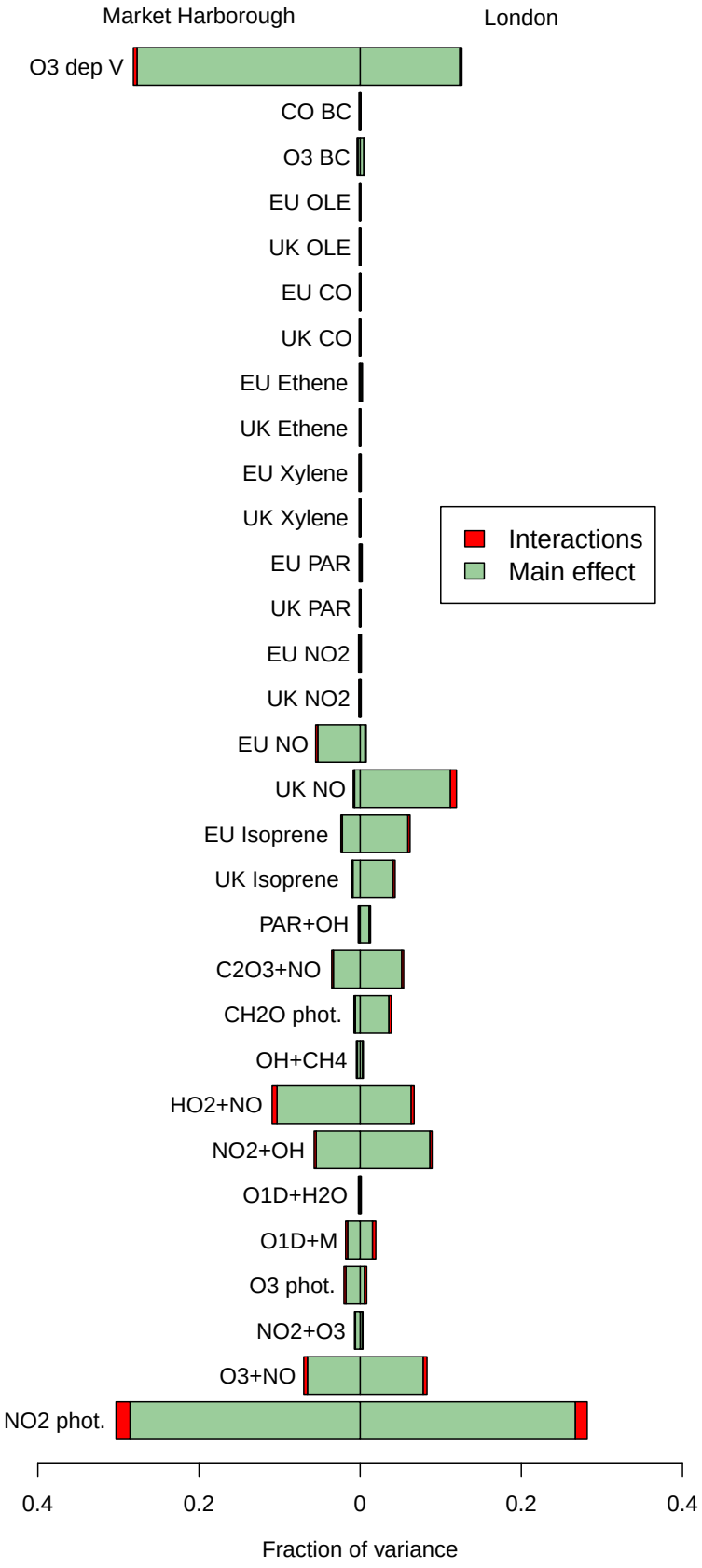


Figure 4.13: FAST for peak afternoon ozone at Market Harborough and London on 19th July.

4.2.4 Nitrogen Dioxide

The three plots in this section show the FAST time series for CMAQ NO_2 output at the same locations as for the time series in the previous section. In all cases the analyses are dominated by UK NO emissions and appear less interesting than the ozone results, however there are some points worthy of note. Again, the UK average plot, figure 4.14, is smoother than the plots for specific locations, particularly in the diurnal pattern of the influence of reaction rate uncertainties. These uncertainties are more influential in the UK plot, perhaps because the inclusion of more remote areas far from major NO_x emission sources means that on average NO_2 has undergone more chemical transformations, particularly in its cyclical relationship with NO and ozone. Directly related to this is the small influence of uncertainty in directly emitted NO_2, evident on all three plots, but smallest on the UK plot and largest on the London plot, figure 4.16. This higher sensitivity in London is probably due to higher levels of directly emitted NO_2 from the vehicle fleet. In common with the ozone plot for London in figure 4.10, the London NO_2 plot shows more interactions than the plots for the other areas, but this time with a less strict temporal pattern. There are some similarities, however, with both plots showing peaks in interactions shortly before midnight on the 11th and 12th. These peaks are associated with sharp peaks in the influence of reaction rates, again possibly related to the transition from daytime to night time chemical regimes.

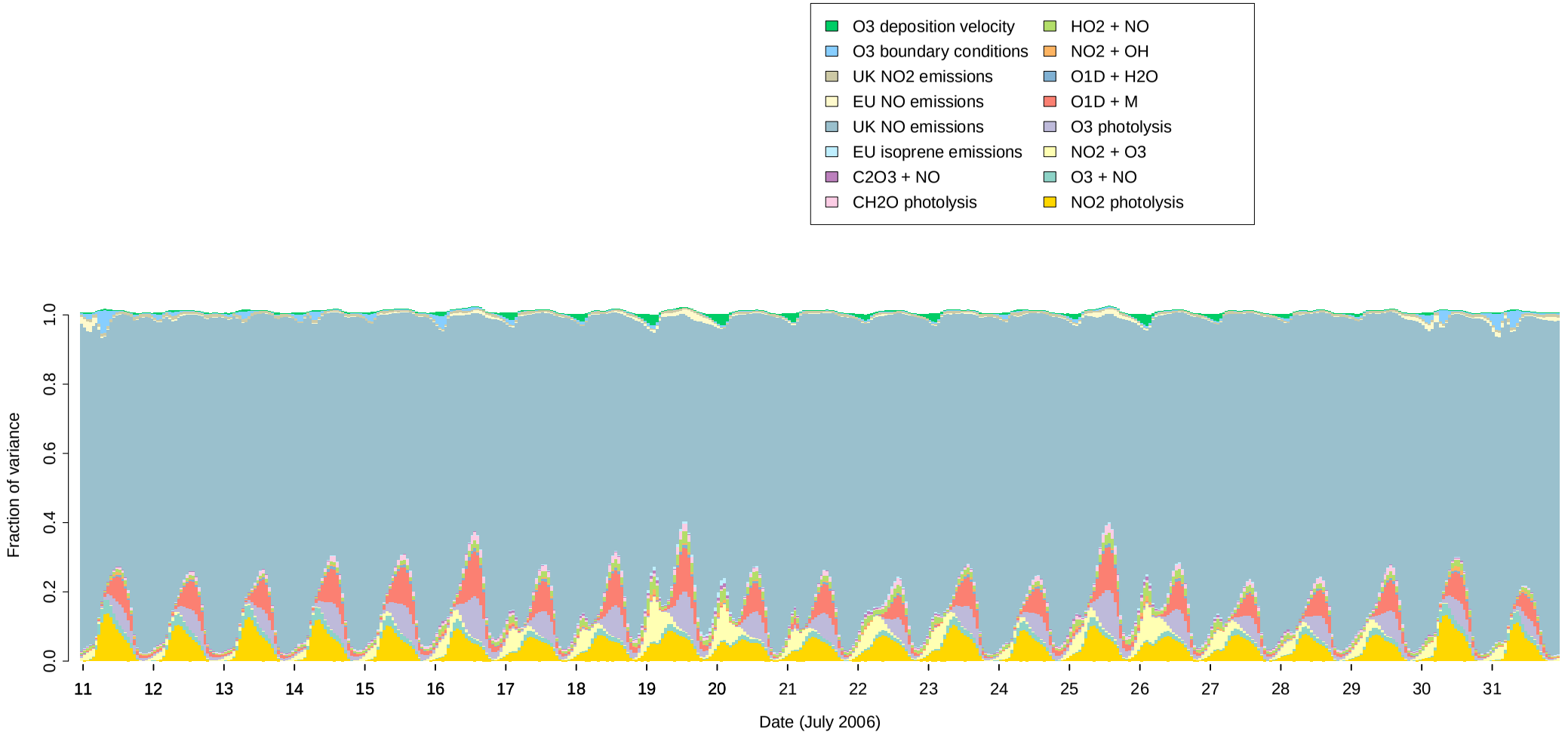


Figure 4.14: Time series of total effects on modelled UK average NO_2 concentrations, 11th-31st July

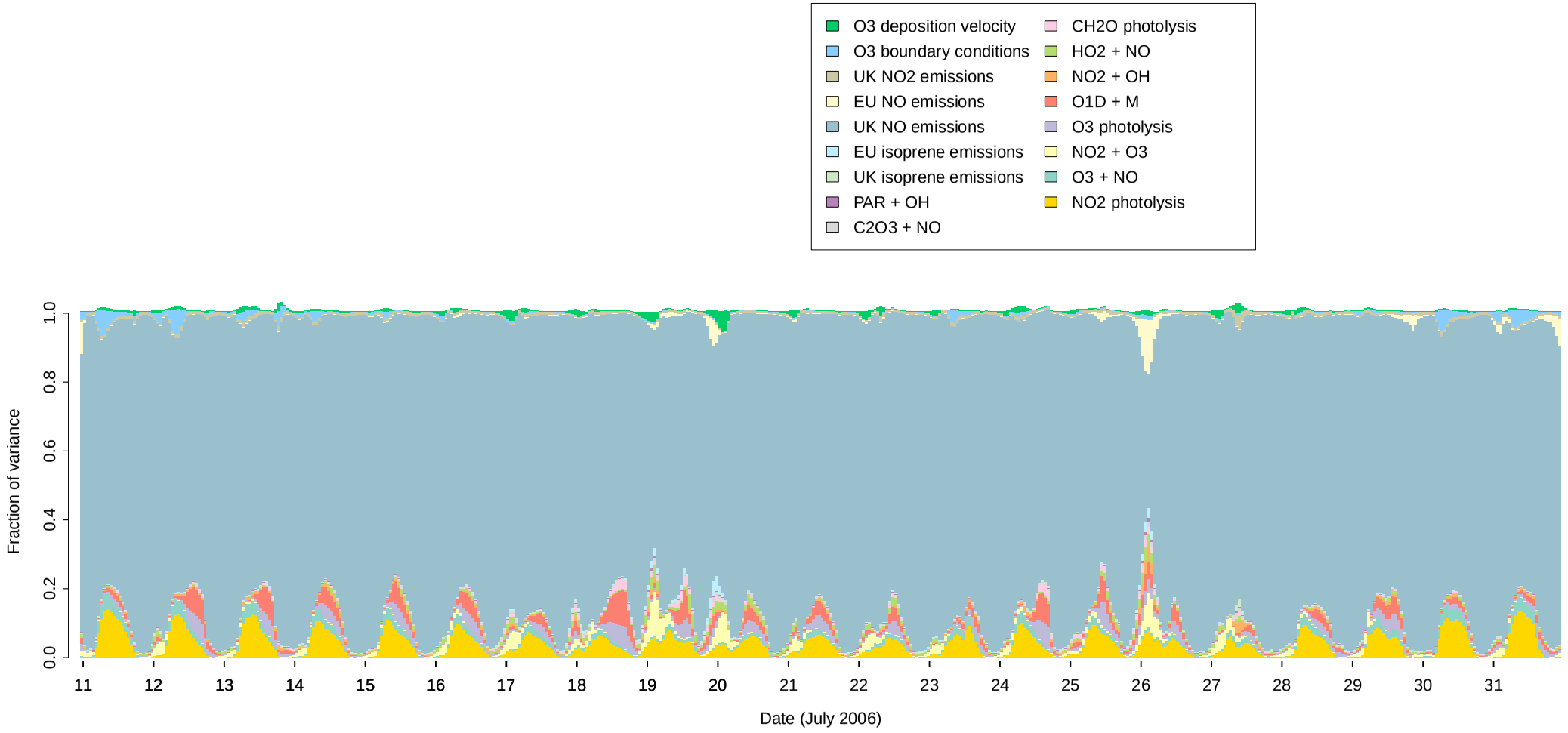


Figure 4.15: Time series of total effects on modelled NO_2 concentrations at Harwell, 11th-31st July

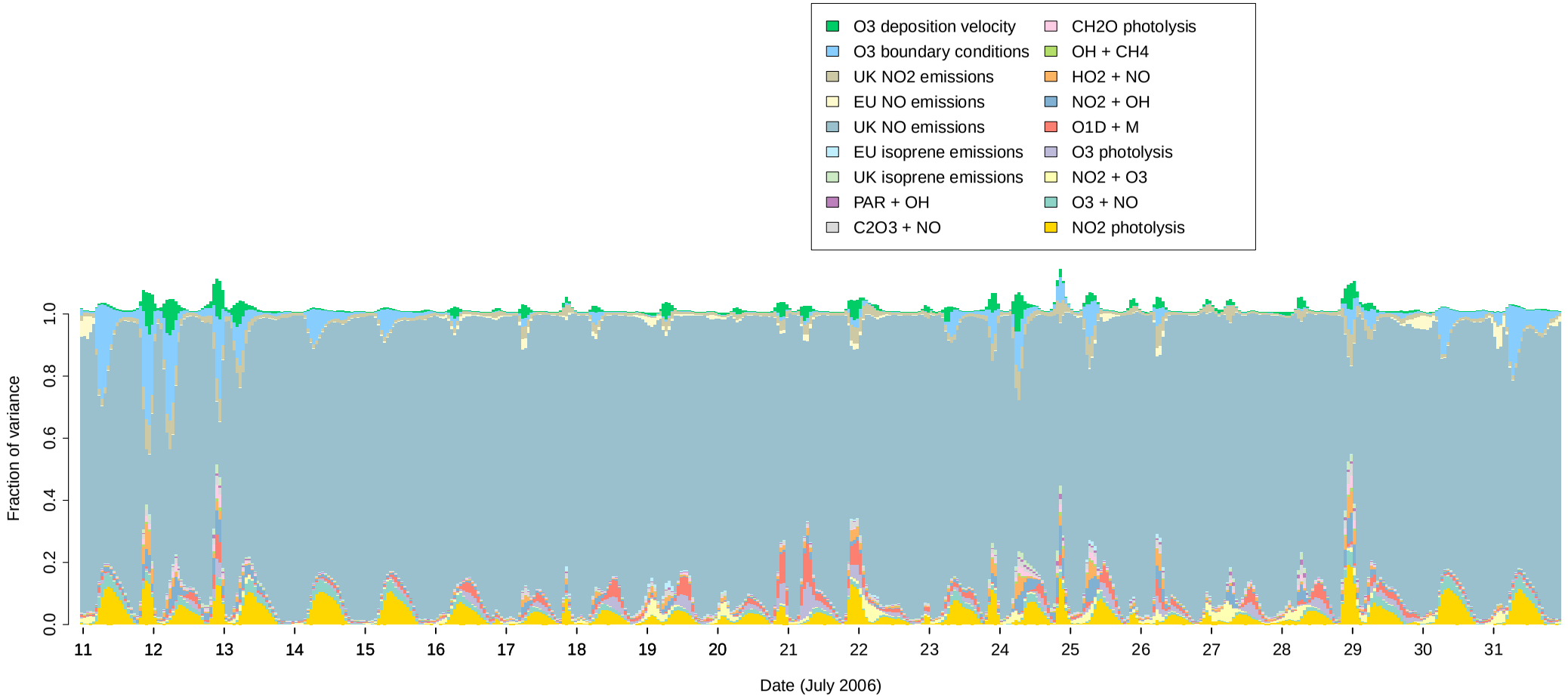
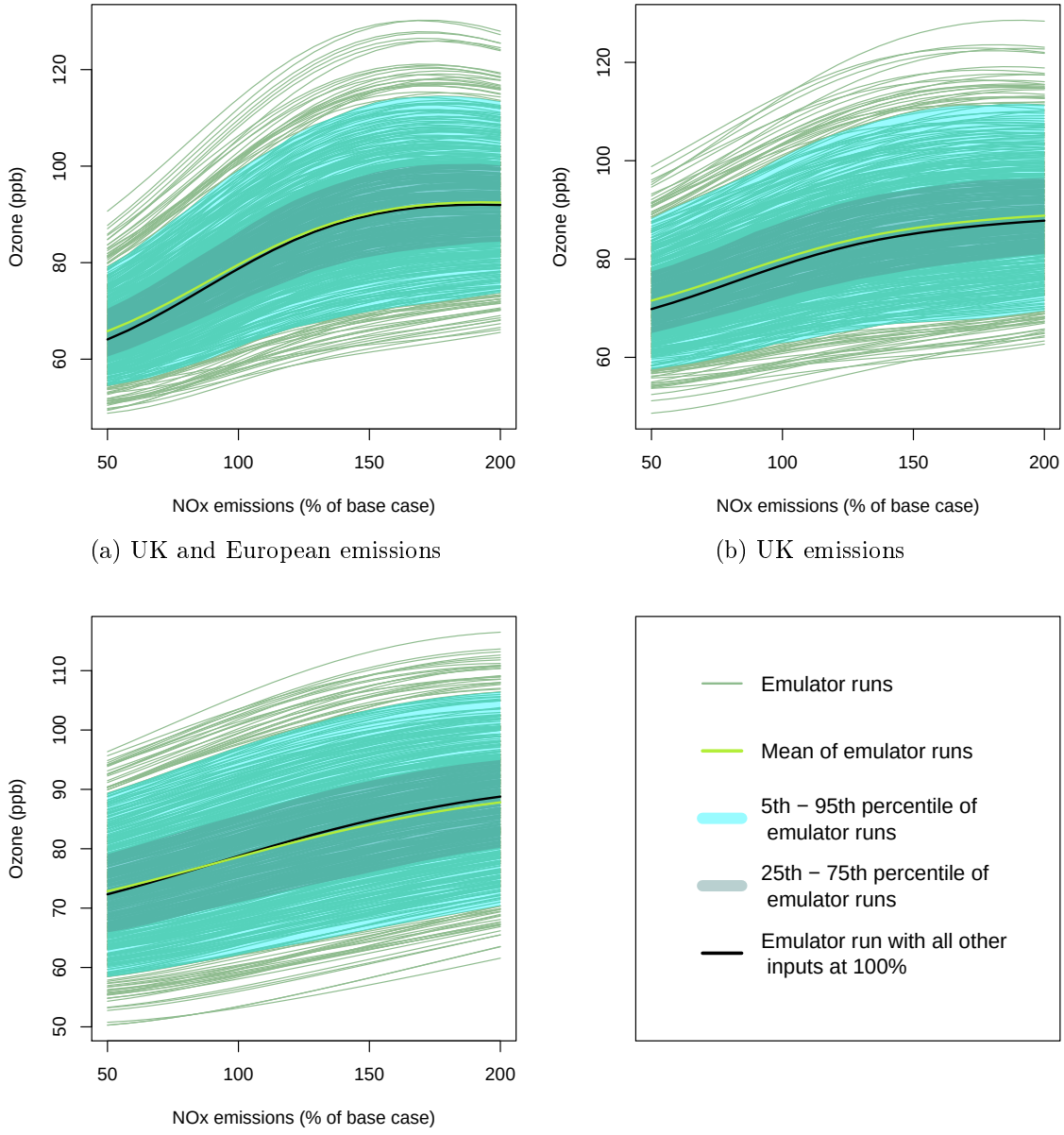


Figure 4.16: Time series of total effects on modelled NO_2 concentrations in London, 11th-31st July

4.3 Accounting for Uncertainty in Main Effects

The plots in this section (figure 4.17, for example) show the effect of varying one input factor at a time on modelled ozone concentrations, and how that effect is influenced by uncertainty in the other input factors. The black line shows the effect of varying the input factor in question over a range of one half to double its baseline value, whilst holding all of the other input factors constant at their baseline values. The darker green lines show the effect of varying the same factor, again with all other factors held constant, but this time at values sampled randomly from their uncertainty distributions. There are 500 such lines on any individual plot, each produced using a different combination of values of the uncertain inputs sampled from their joint distribution. The bright green line shows the mean of the darker green lines and the shaded regions indicate their 5th, 25th, 75th and 95th percentiles. All of the plots were produced with the emulators of the CMAQ ozone output at 4 pm on the 19th July.

Figure 4.17 shows the effect of increasing NO_x emissions on ozone concentrations at the rural Harwell site. Both UK and EU emissions contribute to ozone formation, and their effects are additive, as seen in the top left plot. This effect is reversed, however, in London, shown in figure 4.18. Here the effect on ozone concentration is dominated by UK emissions, which serve to reduce ozone levels via the reaction $NO + O_3 \longrightarrow NO_2 + O_2$. This effect is common at locations close to major NO sources (Sillman and He, 2002). The fact that all of the individual emulator runs produce output which is almost parallel in many cases, shows that if local sensitivity measures were to be calculated they would be fairly robust in the face of uncertainty in other model inputs. In other words, this uncertainty affects the vertical position of the lines on the plot, but the gradients of the lines at any fixed NO_x emission value are fairly similar. However, the danger of extrapolating such sensitivity measures, as has often been done with CMAQ reduced form models and was discussed in chapter 1, can be seen as the rate and direction of curvature on many of the plots changes quite dramatically.



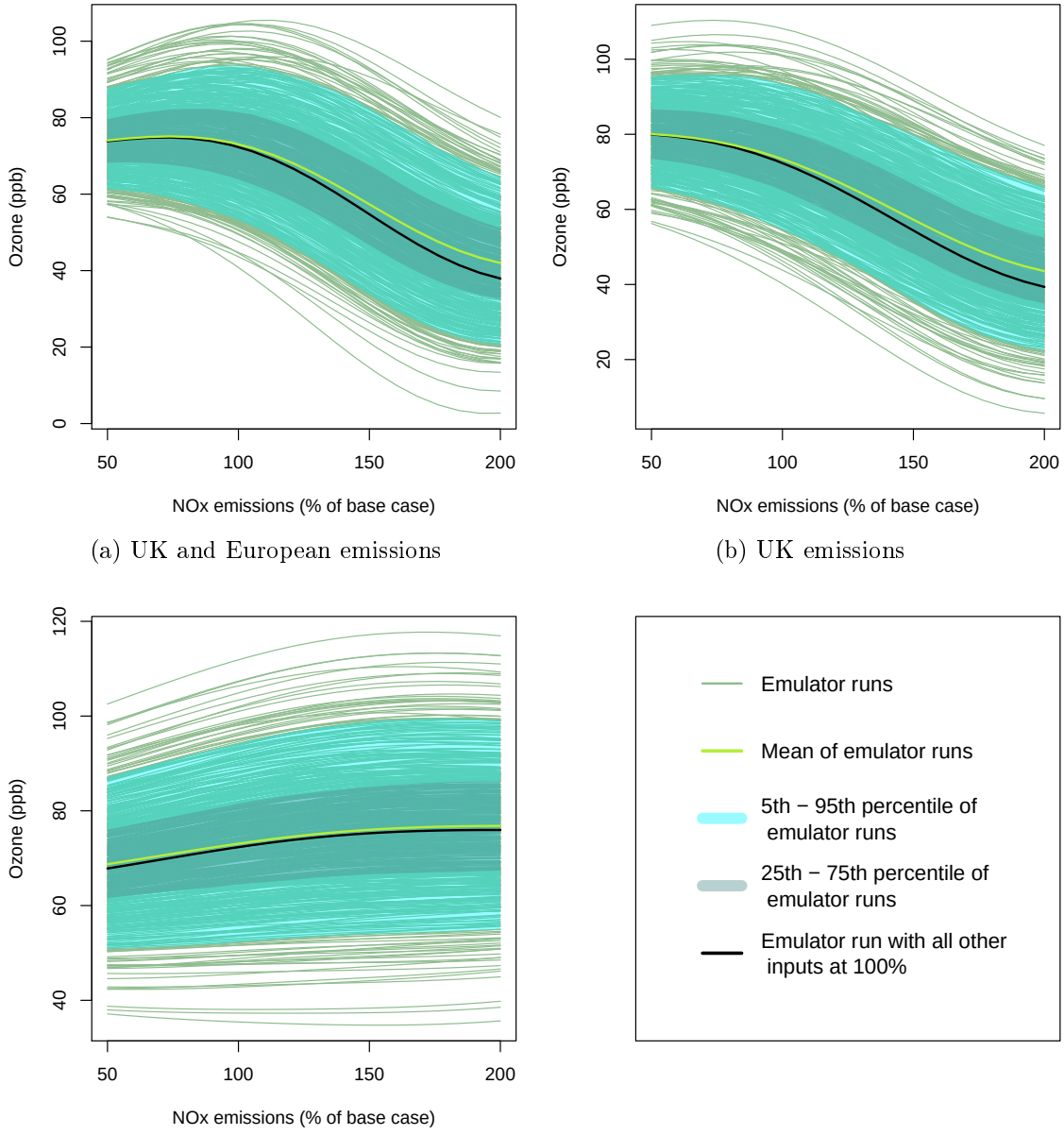
(a) UK and European emissions

(b) UK emissions

(c) European emissions

(d) Legend

Figure 4.17: Sensitivity of modelled ozone to NO_x emissions whilst accounting for uncertainty in other inputs, Harwell, 4pm on 19th July.



(a) UK and European emissions

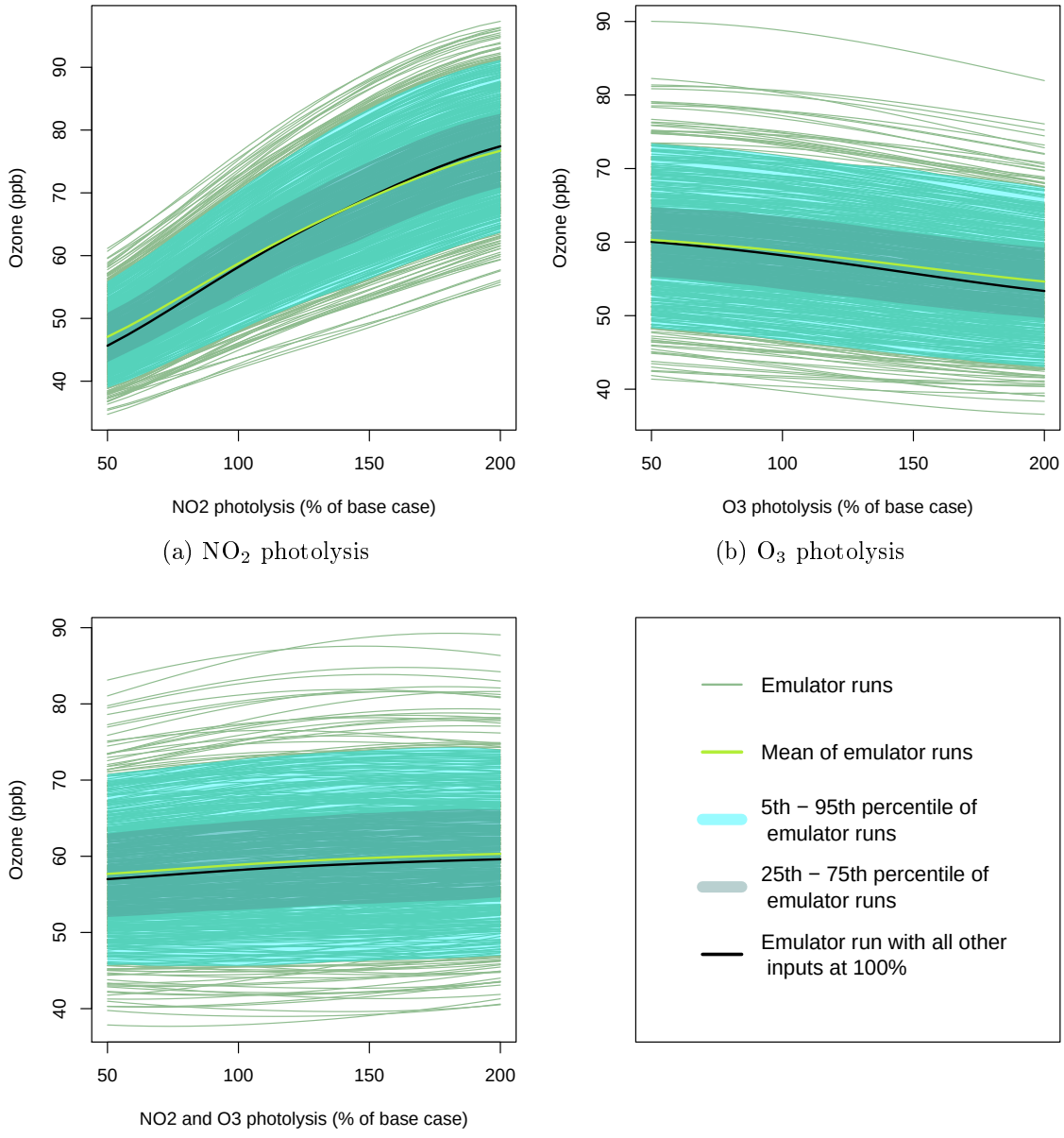
(b) UK emissions

(c) European emissions

(d) Legend

Figure 4.18: Sensitivity of modelled ozone to NO_x emissions whilst accounting for uncertainty in other inputs, London, 4pm on 19th July.

The main effects of NO_2 photolysis, ozone photolysis, and of the two combined are shown in figure 4.19, this time on mean ozone concentrations across the UK. Increasing NO_2 photolysis has a positive effect on ozone concentrations, while increasing ozone photolysis has a negative effect, as would be expected from the chemistry of ozone production described in chapter 1. The relative gradients of these two effects reflects the fact that NO_2 photolysis proved to be considerably more important than ozone photolysis in the variance decomposition of the FAST. Despite this difference in contribution to the overall variabilty of the model output, when the two photolysis rates are varied together, shown in plot (c) of figure 4.19, the effects very nearly cancel each other out. This is an important result, as without this the high degree of sensitivity to the NO_2 photolysis rate shown in many of the analyses would mean that modelled ozone concentrations would also be very sensitive to errors in the modelled actinic flux; that is, to errors in the amount of solar radiation passing through the surface of a unit volume of the atmosphere. This in turn could be caused by errors in the concentration and thickness of the stratospheric O_3 layer, or errors in modelled clouds and particulate matter concentrations, to name just a few of the elements of the model which interract with solar radiation.



(a) NO_2 photolysis

(b) O_3 photolysis

(c) Combined photolysis

(d) Legend

Figure 4.19: Sensitivity of modelled UK mean ozone concentration at 4pm on 19th July to NO_2 and O_3 photolysis whilst accounting for uncertainty in other inputs.

4.4 Bivariate Sensitivity to NO_x and VOCs

The plots in this section are again produced using the emulators for CMAQ ozone output at 4 pm on 19th July, and are similar to the ozone isopleth in figure 1.1, which was introduced during the discussion of NO_x sensitive and VOC sensitive chemical regimes. As mentioned previously, isoprene emissions featured in the results of the sensitivity analyses, but anthropogenic VOCs did not. Producing separate isopleth plots for NO_x vs. isoprene and NO_x vs. anthropogenic VOCs provides a good means of investigating this further.

Figure 4.20 shows the ozone concentrations which CMAQ would have simulated at Harwell for NO_x and isoprene emissions between 50 and 200 % of their baseline values. The intersection of the dashed lines on the plot, representing the base case run, sits on the ridge line between NO_x sensitive and VOC sensitive regimes.

If anthropogenic VOCs are varied instead of isoprene, the plot shown in figure 4.21 is obtained. In this case there is only a small area of the plot, in the top left corner, which could be described as VOC sensitive - reducing anthropogenic VOC emissions would only reduce ozone levels in the case of NO_x emissions being almost double baseline values and VOC emissions being already lower. The rest of the plot is NO_x sensitive, which corresponds with the assertion of Sillman (1999) that ozone levels are more sensitive to NO_x reductions in high biogenic VOC emissions scenarios, of which this is an example, being highly vegetated with afternoon temperatures exceeding 30°C.

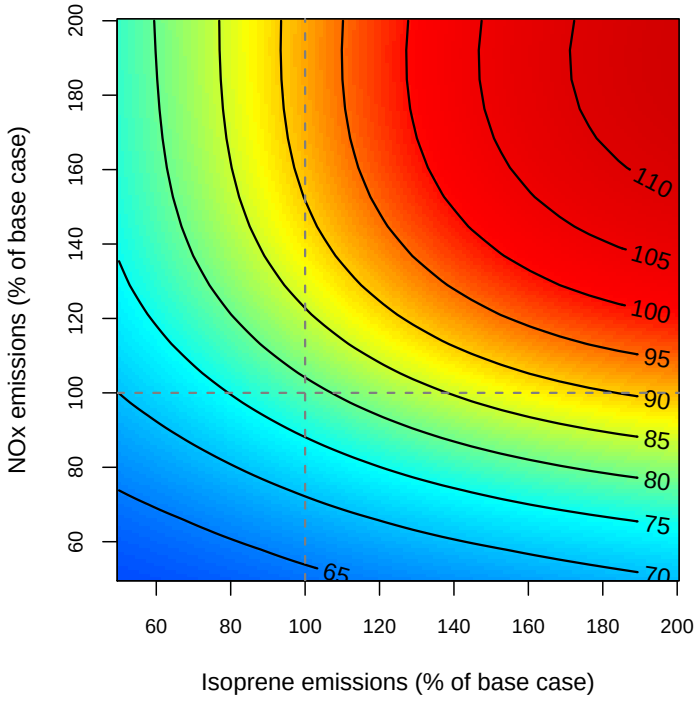


Figure 4.20: Ozone concentrations at Harwell, labelled in ppb on the contours, against NO_x and isoprene emissions.

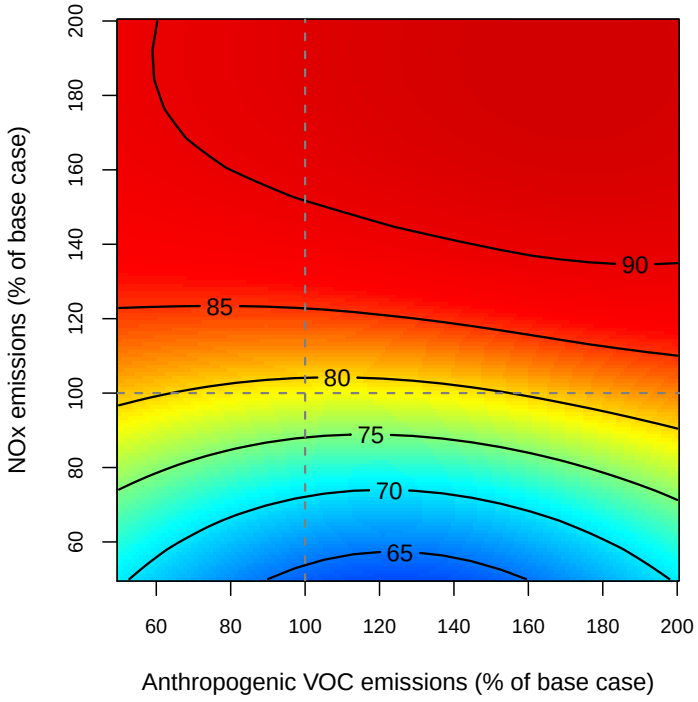


Figure 4.21: Ozone concentrations at Harwell, labelled in ppb on the contours, against NO_x and anthropogenic VOC emissions.

Producing the same two plots for London, the first, shown in figure 4.22, shows that ozone remains sensitive to changes in isoprene emissions across nearly the full range of NO_x emission values, with perhaps just a small area of low NO_x and very high isoprene were VOC sensitivity starts to trail off. Following the dashed line representing the baseline value of isoprene emissions, ozone decreases as NO_x emissions increase from 100 to 200 % of their baseline value. This is the urban ozone titration effect already seen in the main effect plot of figure 4.18.

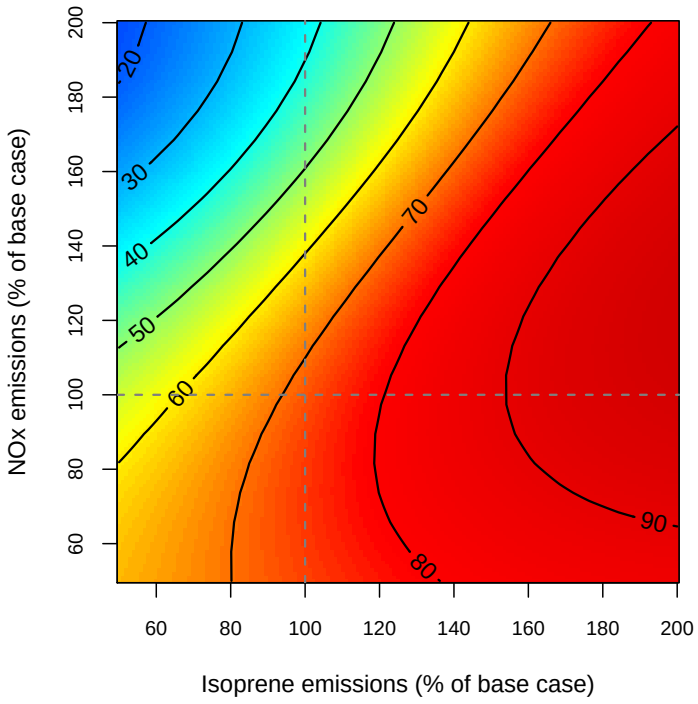


Figure 4.22: Ozone concentrations in London, labelled in ppb on the contours, against NO_x and isoprene emissions.

The titration effect can also be seen in the plot of NO_x emissions vs. anthropogenic VOC emissions for London, shown in figure 4.23. The anthropogenic VOCs appear to be less influential than isoprene, with the contour lines swinging round to be more perpendicular to the NO_x axis in the top half of the plot than in figure 4.22. The anthropogenic VOCs are more influential, however, here in London than in the same plot for Harwell, with the contour lines showing that if NO_x emissions remained constant, then allowing anthropogenic VOCs to increase would increase ozone levels.

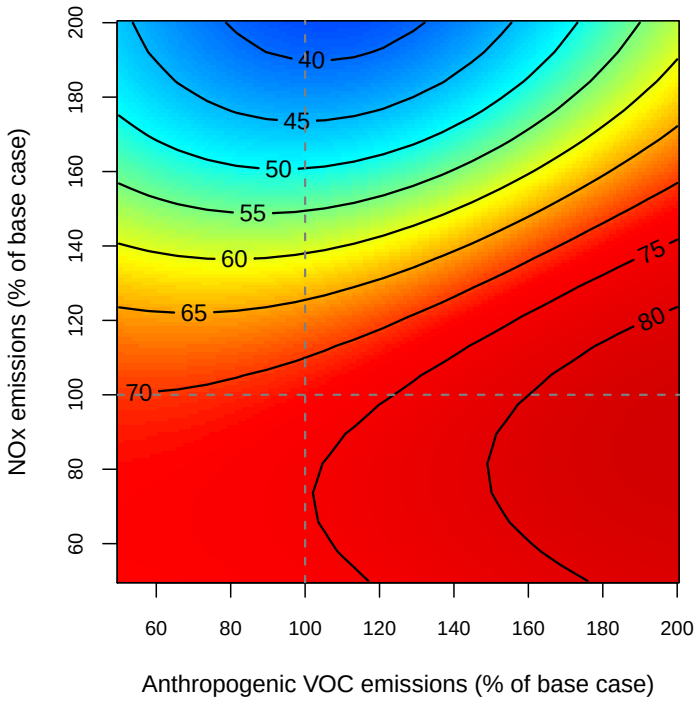


Figure 4.23: Ozone concentrations in London, labelled in ppb on the contours, against NO_x and anthropogenic VOC emissions.

4.5 Final Remarks

The results presented in this chapter are a small subset of those which can be produced having completed the necessary emulator training runs. Concentrating on UK mean, Harwell and London plots hopefully allows the results to be interpreted without swamping the reader with information. Results for additional locations are given in appendix A. The 'take home message' is perhaps that once the active subset of input variables has been identified during the screening process, the the relative sensitivity of CMAQ output to uncertainty in those inputs is highly variable according to time and location.

Chapter 5

Monte Carlo Analysis

This chapter demonstrates the use of emulators to propagate input uncertainty through to the model output in order to produce a distribution of possible values around the single valued deterministic output. This Monte Carlo uncertainty quantification has been made possible with the use of Gaussian process emulation - such an analysis has only previously been achieved with CMAQ using the much less accurate reduced form models described in chapter 1. Two methods, Bayesian Monte Carlo (BMC) and history matching, are then used to incorporate observational data to modify this distribution, and the distributions of the uncertain inputs. As with the main effects plots and the bivariate plots in the previous chapter, the same emulator training data was used as for the sensitivity analysis, so that no new CMAQ runs were required to produce the results presented here.

5.1 Uncertainty Propagation

Here a probability distribution is produced for the model output at a single point in space and time, namely 4 p.m. on the 19th July at Harwell. The same input uncertainty distributions were used as for the sensitivity analysis, so that 31 model inputs had a mixture of normal and uniform distributions, as described in table 4.1. These distributions were assumed to be independent and 100,000 samples were taken from them at random, giving 100,000 different combinations

of input variable values. The emulator was then used to give an estimate of the corresponding CMAQ ozone output at each of these combinations. The resulting distribution is shown in blue in figure 5.1, along with a dashed line in black, which shows the emulator output at the baseline input values. When rounded to the nearest ppb, this is the same value of 79 ppb as the CMAQ base case output, providing reassurance in the accuracy of the emulator. Also shown, with a green dashed line, is the observed ozone concentration of 97 ppb.

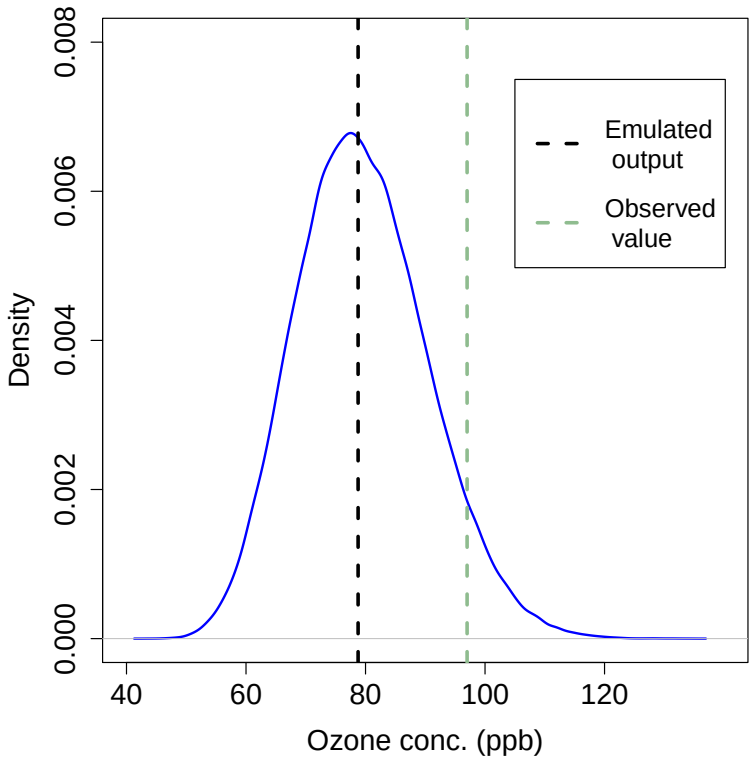


Figure 5.1: Distribution of 100,000 Monte Carlo samples at Harwell on 19th July

An uncertainty distribution produced in this way only represents the output uncertainty induced by uncertainty in the inputs from which the Monte Carlo sample was drawn. Other uncertainties which have not been considered include those inputs ruled out during the screening process, which although insignificant individually may have a noticeable cumulative effect, and inputs which have not been considered at all, such as potential errors in the meteorological inputs. Also, no account of model discrepancy is made in this kind of uncertainty propagation. The distribution in figure 5.1 could easily be summarised by giving confidence

intervals around the CMAQ output, but as with the sensitivity analyses in the previous chapter, these confidence intervals would only be relevant to this particular time and location.

5.2 Bayesian Monte Carlo

The uncertainty propagation in the previous section could be said to be a success, as the observed value is well within the probable bounds around the model output. If this were repeated a number of times (for different times and locations for example) and similar results were obtained, perhaps with a large enough sample to apparently verify the 95 % confidence interval of the distribution, the analyst might be tempted to assert that all significant input uncertainties had been accounted for. However, it may be the case that unaccounted for input uncertainties exist, or that the model discrepancy is significant, and this has been concealed by assigning too much uncertainty to those inputs which have been included in the analysis. In the case of complex models such as CMAQ, which represent the culmination of several decades of development by large numbers of individuals, parameters such as physical constants which are hard coded into the model become buried in the source code and forgotten about over time. If uncertainty exists as to the true value of such constants, this will not be included in an uncertainty analysis, and as a consequence the distinction between model discrepancy and input uncertainty can become blurred.

Even if the above problems do not arise, such an analysis may indicate that the model is of limited practical use if the uncertainty distribution around its output is quite wide; observations may fall within this distribution when subjectively a model user might consider the output to be unacceptably poor. This is where calibration with methods such as BMC may be employed in an attempt to reduce both the input and output uncertainties, and at the same time pull the model output closer to observations.

BMC was applied in order to weight each of the emulator runs made in the

previous section by its posterior probability of being a true representation of reality when compared to observational data. The calibrated output distribution, shown in red in figure 5.2, was produced as defined in chapter 2, using a normal likelihood function with a standard deviation of 2.5 ppb assigned to the observational datum. The desired effect of shifting the output closer to the observations, and narrowing the distribution has been achieved in a somewhat extreme fashion. This is because the standard deviation of the observational error is much smaller than the standard deviation of the uncalibrated (prior) Monte Carlo distribution, and Bayes theorem gives weight to the observational data versus the prior in inverse proportion to these standard deviations.

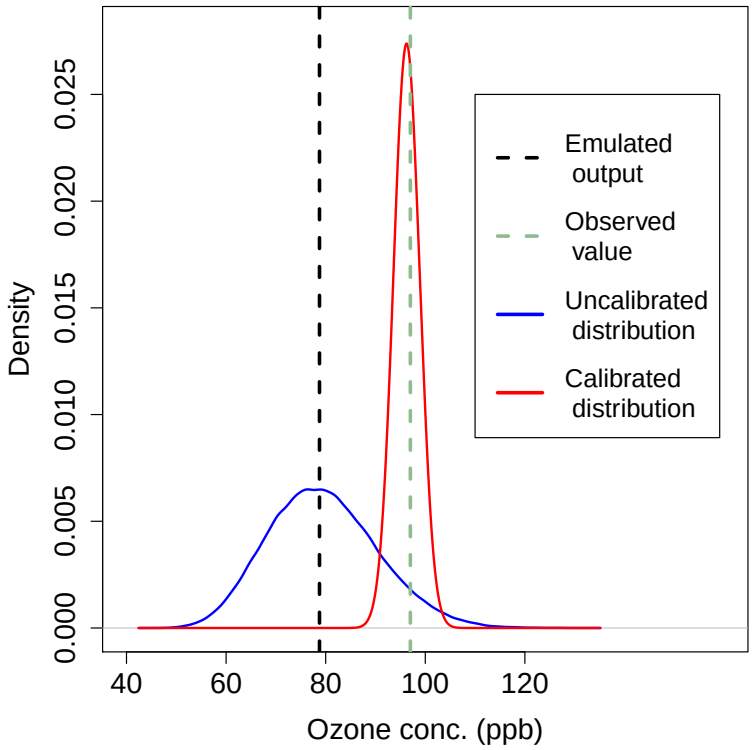


Figure 5.2: Uncalibrated and calibrated distributions of 100,000 Monte Carlo samples at Harwell on 19th July

The weighted Monte Carlo sample allowed modified input variable distributions to be plotted, as each member of the sample represents a different combination of input values, and the six inputs which were most transformed by the process are shown in figure 5.3. Those inputs shown in the figure with uniform

distributions appear as if the assigned uncertainty ranges may not have been wide enough. These ranges were widened and the process repeated again, this time with a sample size of 500,000, and the corresponding input distributions shown in figure 5.4.

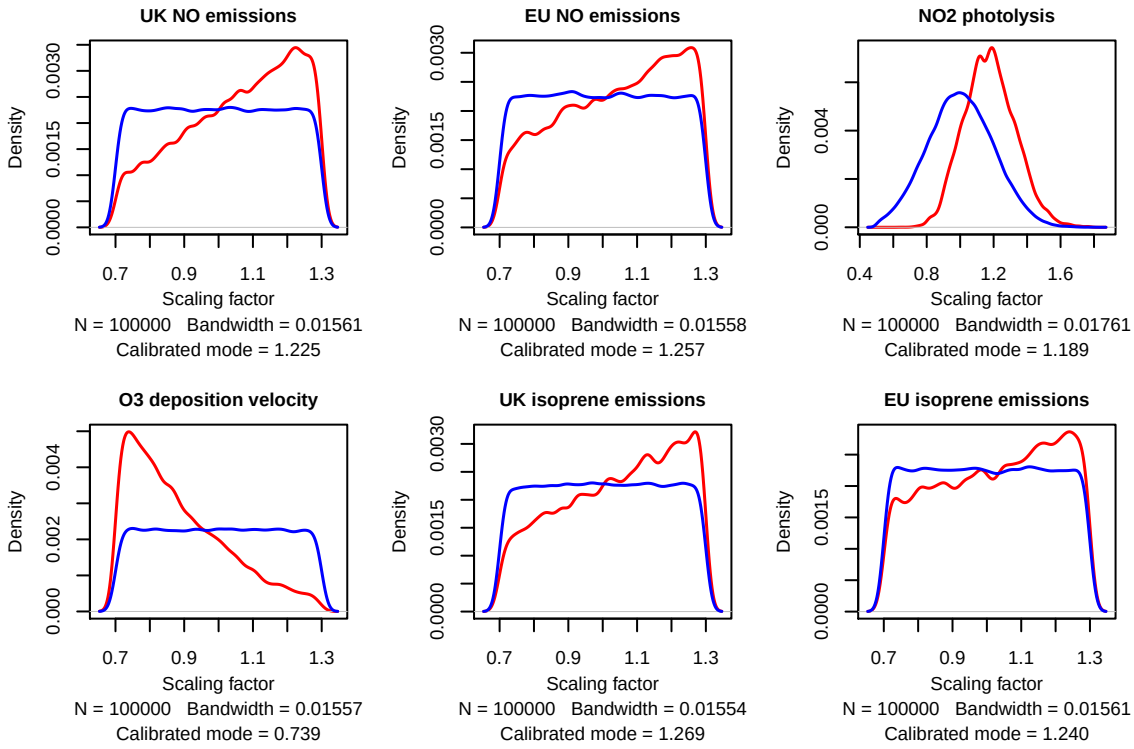


Figure 5.3: Uncalibrated (blue) and calibrated (red) input distributions of 100,000 Monte Carlo samples at Harwell on 19th July.

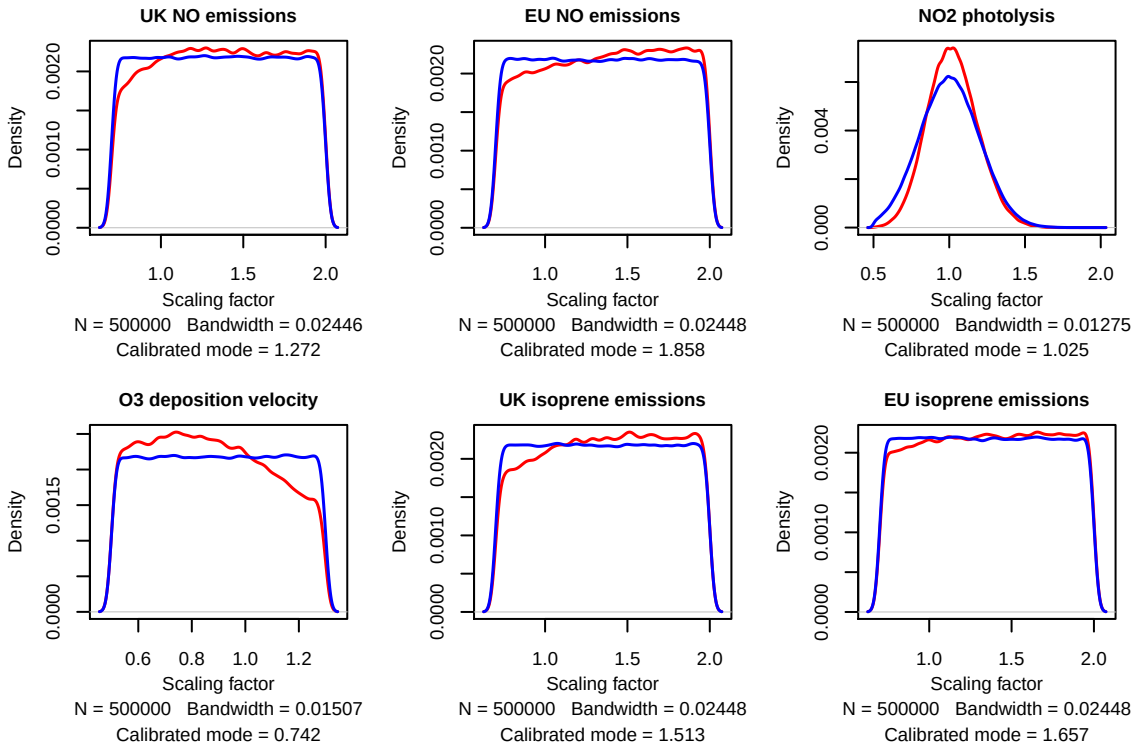


Figure 5.4: Uncalibrated (blue) and calibrated (red) input distributions of 500,000 Monte Carlo samples at Harwell on 19th July, after extending some of the input ranges.

What is immediately apparent is that these wider distributions are in general now much less peaked, indicating that there are many different combinations of input values within these extended ranges that can give output close to the observed value, particularly in terms of the emissions variables. This is an example of the equifinality problem discussed in section 1.7.3. The mean of the NO_2 photolysis rate distribution has now returned to its original position, but unfortunately it is not possible to tell whether this is correct or that a photolysis rate which is too low, therefore producing insufficient NO, has been corrected for by increasing NO emissions. The same calibration exercise was repeated for three other afternoon ozone peaks on the 15th, 16th and 21st July, which are shown in figure 5.5. Initially it was intended to have points equally spaced in time across this pollution episode, but the shape of the peak in measured ozone concentration on the 17th suggested that these measurements may be unreliable. Even if the measurements are reliable, however, the rapidly fluctuating value would suggest

that the ozone concentration is not particularly homogeneous, so the assumption that the monitoring site is located in such a way as to negate the effect of change of support between model grid square and point monitoring location may be unreliable at this time.

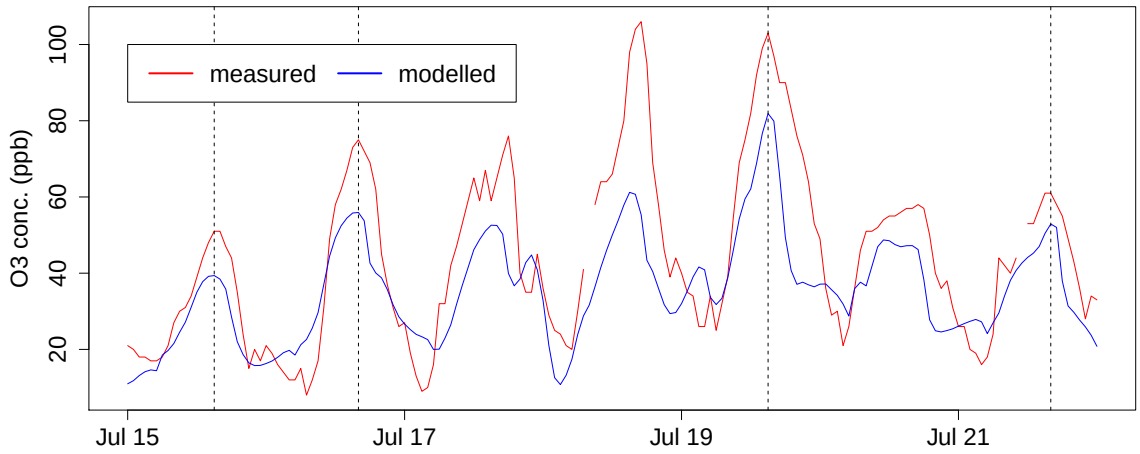


Figure 5.5: Measured and modelled ozone concentrations at Harwell, 15th - 21st July, with dashed lines showing the times BMC was performed.

Each time the method gave different results for the calibrated input variables, an example being given in figure 5.6, which shows how the ozone deposition has been modified differently on each occasion. The same was true for those other inputs which had their distributions modified by the process, and there are three possible reasons for this. Firstly, the method assumes that there is no model discrepancy, i.e. that if the correct values of the input variables were known then the model would reproduce observations exactly. The fact that this is almost certainly not the case probably leads to the method over-calibrating the input variables to make up for the unaccounted for discrepancy. If one considers that this discrepancy may not be a single value but in fact is likely to vary both spatio-temporally and as a function of position in the high dimensional input variable space, then this over-calibration would be different wherever and whenever the method was applied. Secondly, and a point related to the first, is that there are inevitably small errors in the meteorological inputs which have not been accounted for and would also vary on a spatio-temporal basis. Thirdly, the air

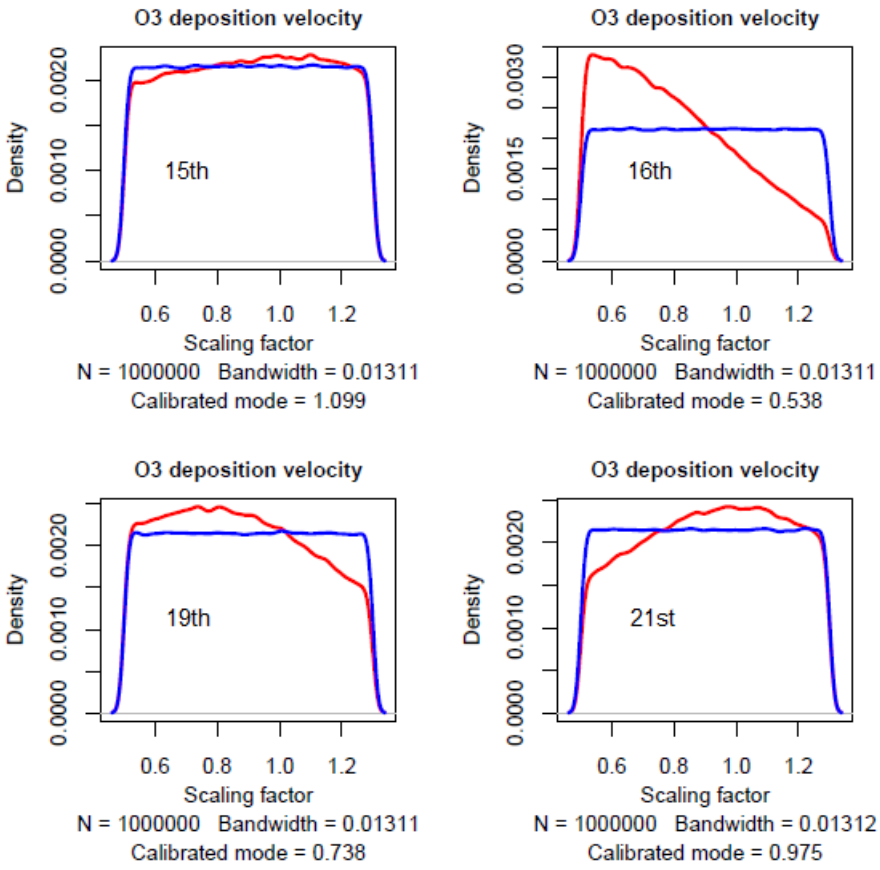
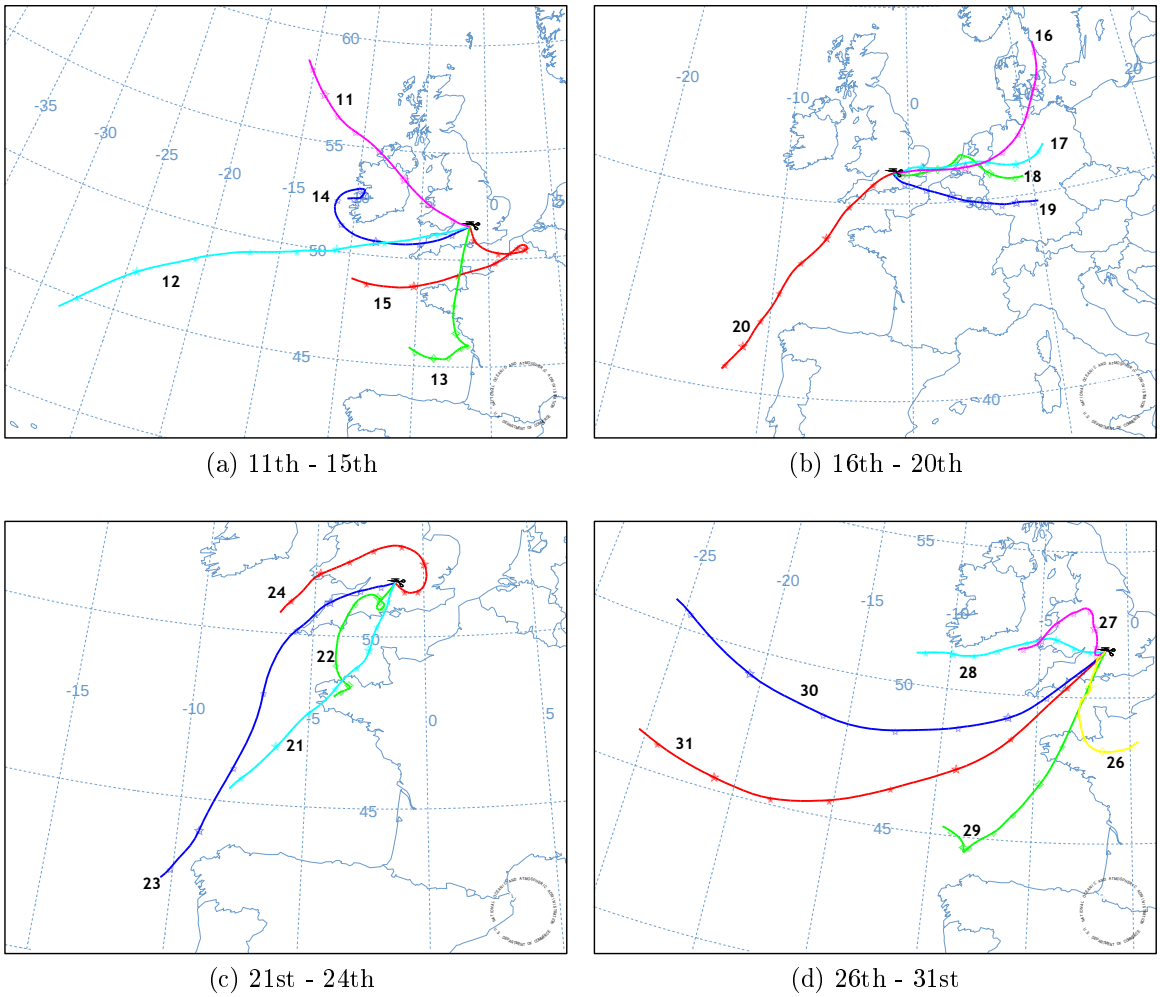


Figure 5.6: Uncalibrated (blue) and calibrated (red) input distributions of ozone deposition Velocity at the afternoon peaks on 15th, 16th, 19th and 21st July.

mass has followed a very different trajectory before arriving at Harwell on each occasion. Figure 5.7 shows the back trajectories of the air arriving at Harwell at 4 p.m. for the whole 21 day period which was used for the sensitivity analyses in the previous chapter, and the days in question are labelled 15, 16, 19 and 21. These trajectories were produced using the online version of the Hybrid Single Particle Lagrangian Integrated Trajectory Model (HYSPLIT-WEB) (NOAA Air Resources Laboratory, 2015). Model inputs which are spatio-temporal fields have been perturbed by multiplying the whole field by a constant factor in all of the analyses in this report, but it is only those grid cells over which the air has passed which will have an influence on the concentration of modelled species, so in all cases except for the most fortunate scenario where a particular input was in error by the same amount everywhere in the model domain, one would expect to find

a range of values for these calibrated inputs.



(a) 11th - 15th

(b) 16th - 20th

(c) 21st - 24th

(d) 26th - 31st

Figure 5.7: 48-hr back trajectories ending at Harwell at 4 pm on 11th - 31st July

In order to investigate this further, the analysis was repeated for every hourly time step in the 21 day period. The ranges of some of the input distributions were extended, as suggested by the previous experiments. Those for both UK and EU emissions of NO, NO_2 and isoprene had their upper limits extended from 1.3 to 2.0 times the baseline value, as did the CO and ozone BCs, and the distribution for ozone deposition velocity had its lower limit extended from 0.7 to .5 times its baseline value. The number of Monte Carlo samples was increased again, this time to one million sample points per time step, giving a total of over half a billion samples drawn from the emulators. In order to make this tractable it was performed in a similar way to the FAST time series in the previous

chapter, splitting the time period into separate days and running each on one of 21 processors. The results for the input variable calibrations were, as expected, extremely varied, producing very different modal values of the distributions at different time steps. The reasons for this have already been outlined above, but nonetheless it was thought that there may be a single value of the scaling factor for each input which would improve the overall model performance over the time period as a whole. It was fully acknowledged that this would not provide correct values for the input variables at each time step, but may give some improvement in model output without performing the much more complex and time consuming task of attempting to calibrate the entire spatio-temporal input fields. A simple average of the modal values for all of the calibrated input distributions was used as the scaling factor for each input. A more conventional method of combining the distributions, such as Bayesian updating using a likelihood function that takes into account the correlation between the time steps, was not felt to be appropriate in this case given that the distributions represent many different locations as the back trajectories swing around across the domain. All of the reaction rate factors had averages close to one, so these were left unchanged, and the rest of the scaling factors are shown in table 5.1.

Table 5.1: Mean values of calibrated input scaling factors

Input variable	**Scaling factor**
UK NO emissions	1.3457
EU NO emissions	1.3828
UK NO2 emissions	1.2692
EU NO2 emissions	1.2904
UK ISOP emissions	1.4044
EU ISOP emissions	1.5104
UK PAR emissions	0.9087
EU PAR emissions	0.8875
UK XYL emissions	0.9331
EU XYL emissions	0.8969
UK ETH emissions	0.9167
EU ETH emissions	0.9230

Table 5.1: (continued)

Input variable	Scaling factor
UK CO emissions	0.9083
EU CO emissions	0.8867
UK OLE emissions	0.9439
EU OLE emissions	0.9014
O3 BCs	1.3444
CO BCs	1.3559
O3 deposition velocity	0.9682

These input factors were used as input to emulators of each time step, as a test to see whether it would be worth going on to use them with the CMAQ model itself. Figure 5.8 shows the measured ozone concentrations along with emulator output using scaling factors of one to represent the uncalibrated model, and emulator output using the factors above to represent the calibrated model. The calibrated model does appear to perform a little better than the uncalibrated model overall, although as expected not everywhere, and at a few points is actually worse.

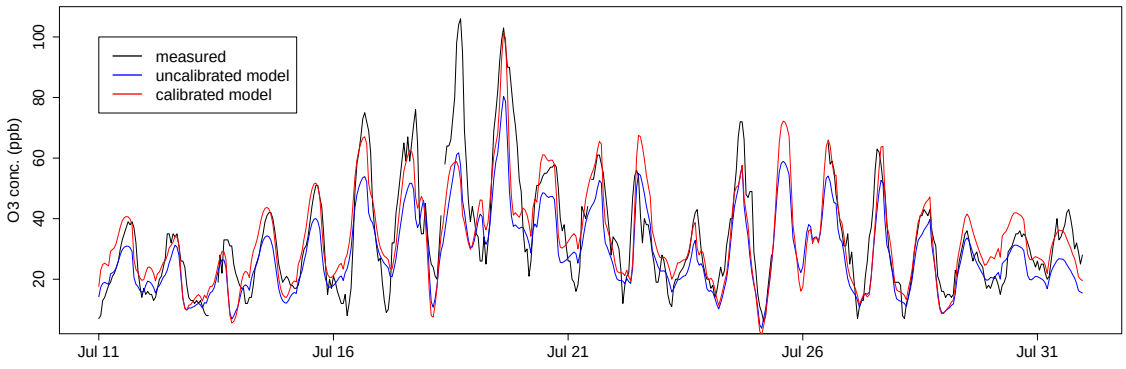


Figure 5.8: calibrated and uncalibrated model output along with measured ozone concentrations at Harwell, Oxfordshire, 11th - 31st July 2006.

Of particular interest are the two prominent peaks representing the afternoon ozone peaks on the 18th and 19th July. The calibrated model captures the top of the peak on the 19th almost perfectly but performs no better than the uncal-

ibrated model on the 18th, suggesting that the model error on the 18th may be largely due to some unaccounted for uncertainties, such as an error in the wind fields which are input to the model. Looking back again to figure 5.7, the afternoon peaks on the 16th, 17th and 18th have similar back trajectories, at least for the preceding 24 hours or so, whereas on the 19th the trajectory has swung around to a more southeasterly direction. The modelled ozone concentrations have a comparable pattern, with the 16th, 17th and 18th having similar sized peaks, but the 19th having a higher peak. The measured concentrations show a different pattern, however, with the 16th and 17th having smaller, equal sized peaks, and the 18th and 19th having larger, but also equal sized peaks. Taken together, this seems to suggest that the modelled wind direction may have swung around more slowly than the real world wind fields.

The same calibrated input scaling factors were used again in another test using emulators of the same time period, but this time for ozone output at a different location. Bottesford, in Leicestershire, was chosen as it was one of the locations used for the variable screening process, has similar characteristics in terms of being a rural background location, and is a large enough distance away, at 155 km, to provide a reasonable test of whether a calibration made at one location is suitable for use at another. The time series of results is plotted in figure 5.9, and the relative locations of Harwell and Bottesford are shown in figure 5.10. It is immediately apparent from this graph that the calibrated scaling factors from Harwell do not work as well at this new location, reinforcing the point made earlier that the correct calibrations are likely to vary both spatially and temporally, rather than being a single value. Also, as the mean calibrated scaling factors were calculated over back trajectories which can be seen in figure 5.7 to come mostly from the west, south and east rather than to the north of Harwell, where Bottesford is located, it is unsurprising that this test produced a smaller improvement in performance at the new location. It is also possible that here the modelled meteorology failed to capture a shallow, stable night-time boundary

layer which might not have been an issue at Harwell.

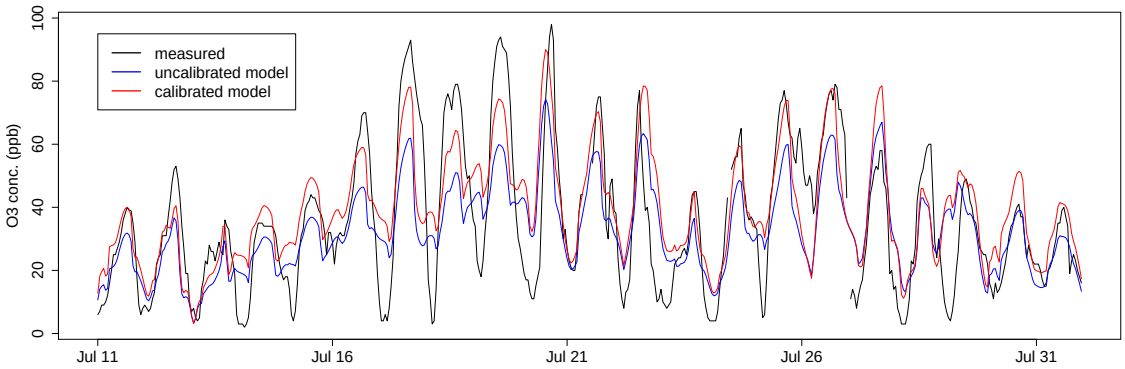


Figure 5.9: calibrated and uncalibrated model output along with measured ozone concentrations at Bottesford, Leicestershire, 11th - 31st July 2006.

The mean bias and mean gross error, calculated using every time step (whole period) and the time steps covering 2 pm to 6 pm on each day (afternoons), for the calibrated and uncalibrated emulators at both locations are given in table 5.2. At Harwell, for the whole period, there is a small improvement in mean gross error, and a large improvement in mean bias, showing that the calibration has not improved the overall performance greatly, but has confronted some of the problems in capturing the afternoon ozone peaks. The same is not true for the whole period at Bottesford, where there was a similar small improvement in mean gross error, but the mean bias shows a shift from under predicting observations to over predicting. Moving on to the afternoon periods, which are when the peak concentrations occur and are therefore of the most interest in terms of health effects and regulatory compliance, the improvements in the calibrated emulator output are considerably greater and are comparable at both locations.

These results demonstrate that a half-way-house approach between this attempt to produce calibration factors which can be applied to a whole spatio-temporal field and the much more demanding task of calibrating each part of those fields individually may be useful in practical applications. It may be possible to produce a set of characteristic calibration factors which could be applied with certain air mass trajectories, weather conditions, time of day etc. This would

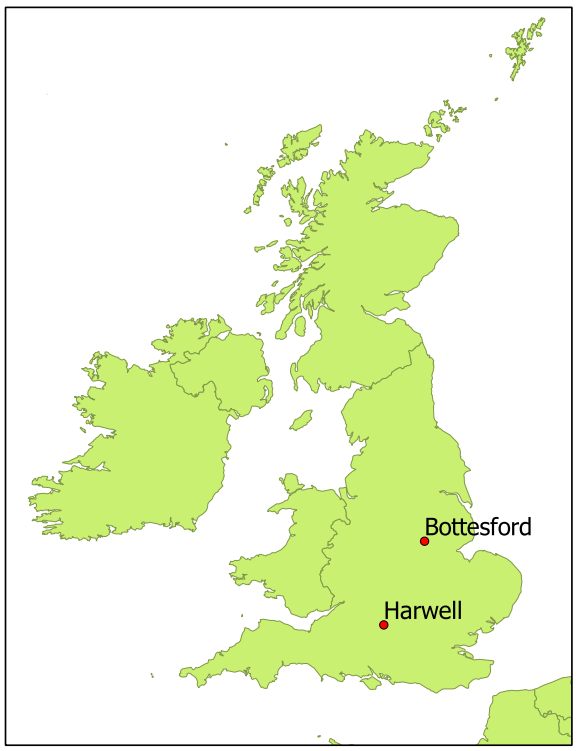


Figure 5.10: Locations of Harwell and Bottesford

Table 5.2: Mean bias (MB) and mean gross error (MGE) of the uncalibrated and calibrated emulator output.

		Whole period		**Afternoons**	
		MB (ppb)	MGE (ppb)	MB (ppb)	MGE (ppb)
Harwell	Uncalibrated	-5.79	8.11	-10.07	11.22
	Calibrated	-0.45	7.48	-1.60	7.28
Bottesford	Uncalibrated	-3.12	12.03	-11.40	14.16
	Calibrated	3.45	11.23	-0.66	9.97

have the advantage that the calibration could be used in circumstances where emulator training runs had not been made, and is a possible future direction for this kind of work.

When the method was described in chapter 2 it was mentioned that using the observational error in the normal likelihood function could be seen as unconventional. This method, as stated earlier in the report, is very similar to the GLUE method. In the critical appraisal of GLUE by Stedinger et al. (2008), the authors assert that GLUE would equate to a correct Bayesian analysis if a correct likelihood function were used. In this case that would equate to using the distri-

bution of model errors for the standard deviation term in the likelihood function rather than the observational error. The effect of this on the results was tested by running the 21 day calibration at Harwell again, this time using the correct standard deviation in calculating the likelihood of each Monte Carlo sample. The output distribution at any point in time was again pulled towards the observational datum and was narrowed, but was now modified less than before, reflecting the fact that the the standard deviation of the model errors, whilst varying from one time step to the next, is generally in the order of 10 ppb whereas the observational standard deviation was only 2.5 ppb. The calibrated input factors, when aggregated across the whole period as before, were almost identical.

The calibrated model output produced by such procedures cannot be said to be probabilistically correct in any case as the model discrepancy and other uncertainties have not been accounted for. A fully Bayesian or Bayes linear analysis which accounted for all sources of uncertainty, even if it were tractable, would be unlikely to produce more useful results without using different or more observational data because of the equifinality in model output which was suggested by the very flat calibrated input distributions in figure 5.4. There are, however, a great deal of available observational data for ozone, NO, NO_2 and a more limited amount of data for certain VOCs, so these preliminary experiments show that designing an experiment to take advantage of such data may yield worthwhile results.

5.3 History Matching

A preliminary attempt at history matching was carried out, again using an emulator trained on the 576 CMAQ runs used in the sensitivity analysis, for peak ozone levels at 4 p.m. on 19th July at Harwell. The history matching process was defined in section 2.5 and aims to reduce the ranges of the input variables by ruling out combinations of inputs that produce model output which is deemed to be implausible according to equation 2.19. A nominal value of 5 ppb standard de-

viation in ozone concentration was used for the model discrepancy, reflecting the fact that the model is capable of performing reasonably well in predicting ozone based on previous experience, but still being double the observational error. This value was chosen with acknowledgement of the fact that a more rigorous estimate would have to be made if the method proved useful for future use. The fact that the BMC in the previous section produced calibrated input distributions which in some cases have maxima right at one end of the allowed range, seems to suggest that the uncertainty distributions assigned to those inputs may be too narrow. As a result history matching was carried out with the full ranges of the input variables over which the emulator training runs had been made, namely one half to double their baseline values. The results of 100,000 Monte Carlo runs are shown in figure 5.11 as a histogram of model output values before the history matching in light blue. Those model runs which were retained as being not implausible are shown superimposed in dark blue.

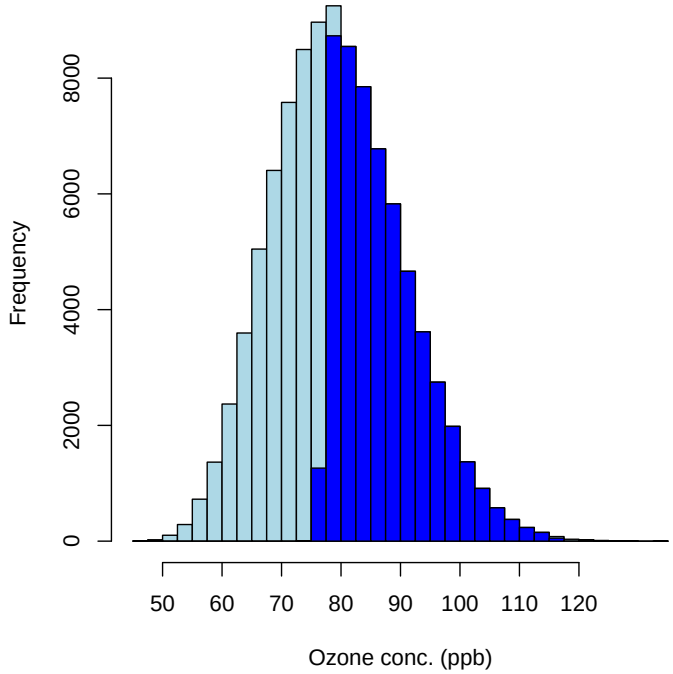


Figure 5.11: 100,000 Monte Carlo samples before history matching (light blue) and 55,676 after (dark blue), Harwell, 19th July.

The range of the not implausible model output can be seen to be roughly

symmetrical around the observational value of 97 ppb, but the modal value of the distribution is not shifted - a consequence of the fact that here model runs are simply ruled in or out and not assigned weights in any way. Despite the fact that over 44 % of the 100,000 Monte Carlo runs were ruled out, none of the marginal distributions of the input variables were constrained by the process. Those distributions are shown in the histograms in figure 5.12, with light blue again showing the samples before history matching which all have flat tops as uniform distributions were used for all of the inputs. Again, those runs remaining as not implausible are shown in dark blue, and it is easy to see that none of the ranges come close to being constrained, except perhaps for NO_2 photolysis, but even here, at the lower end of the distribution still around one third of the runs remain as not implausible.

In order to gain a better insight into how the implausibility was spread across the input variable space, plots were produced which Williamson et al. (2013) describe as 'not ruled out yet (NROY) density plots'. Following the method of Williamson et al., two of the input factors are fixed, the emulator is sampled at 1000 points according to a Latin hypercube sampling plan defined over the remaining 29 dimensions, and the implausibility calculated for each point. The number of emulator 'runs' which are deemed to be not implausible can then be represented as a proportion of the total number of 1000 runs. This was repeated at fixed values of the two inputs of between one half and double their baseline values, allowing the proportion of not implausible runs to be plotted as a surface. This is shown for four different pairs of input factors in figure 5.13. The colour scale on the plots represents the proportion of the Latin hypercube points which are not implausible, so a value of one would indicate that at those particular values of the inputs on the axes, none of the combinations of the remaining input variables could be said to be implausible, and a value of zero would mean that all of the combinations of the remaining inputs are implausible.

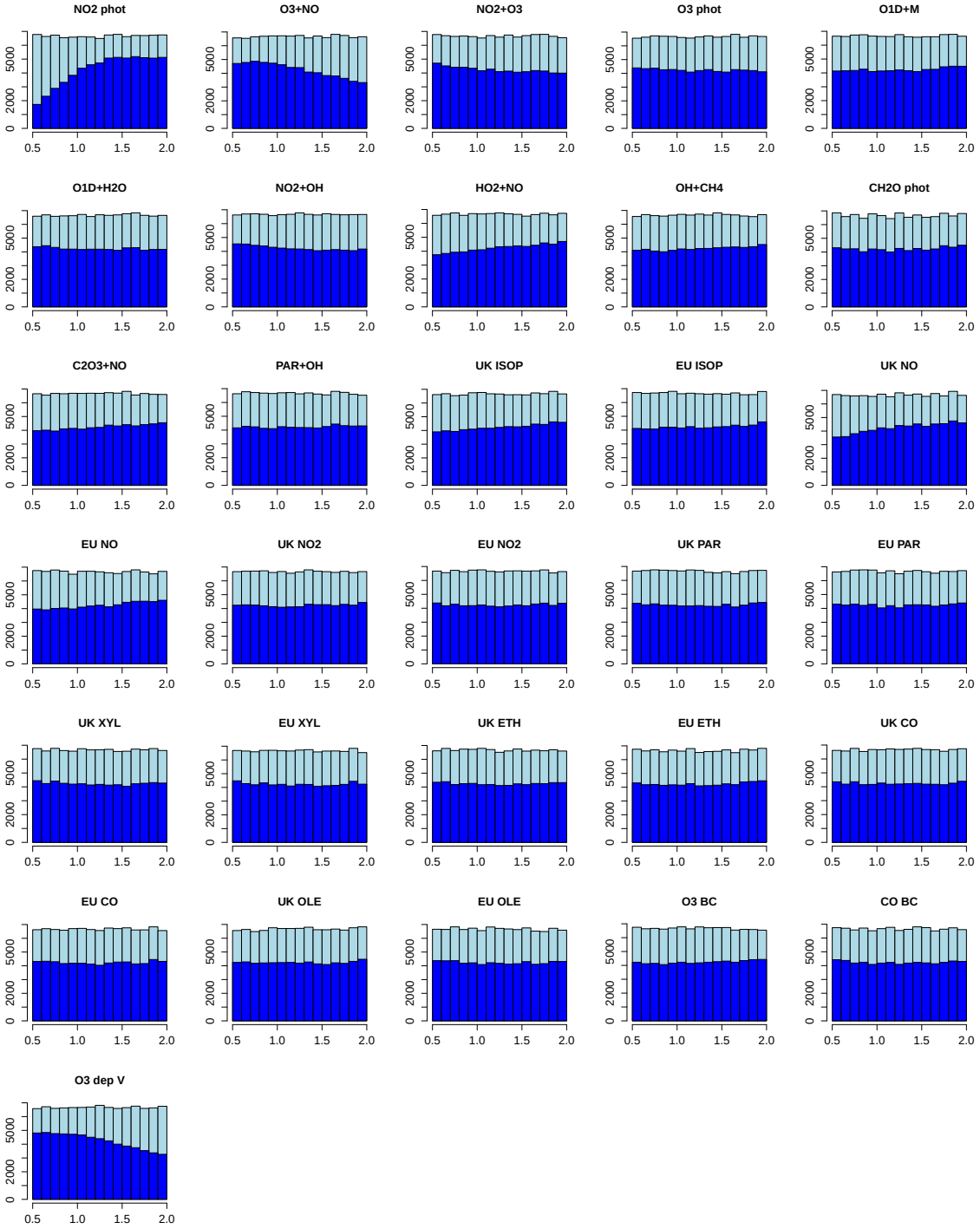


Figure 5.12: Input distributions of 100,000 Monte Carlo samples before history matching (light blue) and 55,676 after (dark blue), Harwell, 19th July.

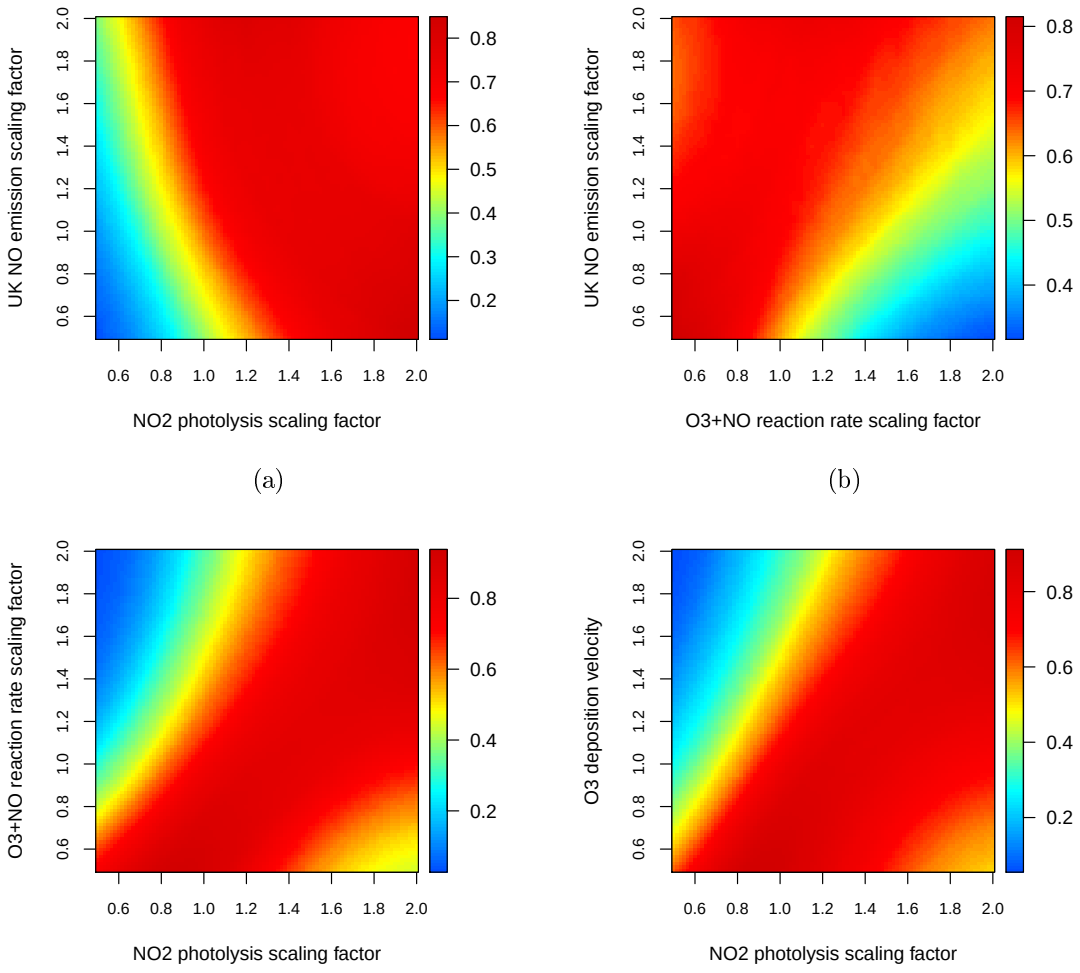


Figure 5.13: Proportion of points which were not implausible when pairs of inputs were held constant at values across their ranges, Harwell, 19th July.

The large areas shaded in red on all four plots in figure 5.13 indicate that there are large areas of the input variable space which contain possible valid values for each of the pairs of inputs when projected onto those two dimensional planes. The blue areas on each plot represent combinations of values of the two inputs which are unlikely to be valid candidates for the true values of those inputs, and as these blue areas comprise fairly small corners of the plots in which we would not expect the true values to lie anyway, it could be said that history matching against this particular measured variable has revealed little about the input variable space. However, if observational data for a second modelled species were available, the implausibility of that emulated model output could also be calculated for each

point in the input space and the maximum value of implausibility with respect to the two observations assigned to that point, as described by Vernon et al. (2010). This would probably reduce the area in the plots in figure 5.13 which is likely to be not implausible, and would hopefully constrain some of the marginal ranges of the input factors. NO_x and isoprene are both measured at this particular site and would both be likely to add useful information because of the complex chemical processes relating ozone, NO_x and VOCs which are represented in the CMAQ model. Unfortunately, this data is unavailable for the time period in question, probably because of equipment failure. The time period studied was initially chosen because the poor model performance at the time of an ozone pollution episode was interesting in terms of performing a sensitivity analysis. However, the availability of observational data on species other than ozone was not considered when making the initial decision on which time period to study.

Vernon et al. (2010) describe an iterative process of performing several waves of history matching, each time re-sampling from the reduced NROY space to construct more accurate emulators, so it would perhaps be worth trying a second wave in this case to see whether any more of the input space could be ruled out.

5.3.1 The Effect of Mis-specifying Discrepancy

The expriment was repeated with different values of standard deviation for the model discrepancy, from zero to ten ppb. The results of this, in terms of the number of runs which were deemed to be implausible, are shown in figure 5.14. The effects of mis-specifying the discrepancy can be seen to be extreme, with a standard deviation of 10 ppb ruling out fewer than 10 % of the model runs, while setting the discrepancy to zero results in nearly 70 % of the runs being deemed implausible. As the variance terms representing the observational error and the emulator accuracy also appear in the denominator of equation 2.19, an increase in any of these terms would cause the number of implausible runs to decrease.

It could well be the case that the model discrepancy in CMAQ does not

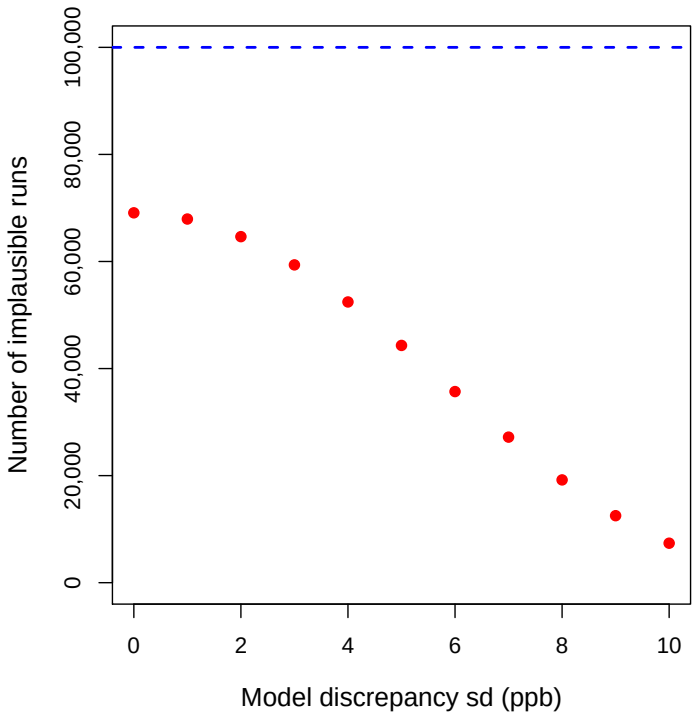


Figure 5.14: Number of implausible runs, out of 100,000, calculated with different values of model discrepancy, Harwell, 19th July

have a fixed value but is a function of one or more variables in the input space and/or the output space. For example, biogenic isoprene emissions have a positive exponential relationship with increasing temperature (Laffineur et al., 2011), so that any errors in isoprene chemistry, as represented in CB-05, would not come into play until air temperatures were high enough, and model discrepancy would then increase exponentially with increasing temperature.

In terms of constraining the input factors, even the case with zero discrepancy, when 69,095 out of the 100,000 runs were ruled out, failed to significantly reduce the marginal ranges of any of the input variables. This is probably a result of the high dimensional input variable space and the complex nature of the interactions in the chemistry represented in the model. Of course, one could conclude that the uncertainty ranges assigned to the inputs were too small to begin with, and this is the reason why none of the marginal distributions were reduced. However with such a large proportion of model runs being ruled out it seems more likely that lack of identifiability is the real reason.

5.4 Final Remarks

The work in this chapter represents a first foray into the area of model calibration with CMAQ, and as such is a snapshot of work in progress. Although the results have shown varying degrees of success so far, they do indicate what the next steps should be. In order to improve the chances of success of any kind of calibration or history matching process, the number of species for which observational data is available which correspond to CMAQ model species should be a prime consideration in choosing the time period over which to attempt the process. In other words, performing calibration as a designed experiment, rather than with data for a period which was primarily chosen as an interesting period for a sensitivity analysis, will have a much greater chance of success. It is also worth noting that nothing which has been attempted here could even be considered without the use of emulators, so it has been demonstrated that this kind of work is at least a possibility.

Chapter 6

Conclusions

The use of Gaussian process emulation has allowed much more detailed sensitivity and uncertainty analyses of the CMAQ model to be undertaken than has previously been possible. The results have demonstrated that using such methods makes it possible to analyse complex air quality models in a robust manner and that doing so only requires moderately powerful computing resources. The sensitivity analyses in particular have revealed that the spatio-temporal pattern of the sensitivity of modelled ozone concentrations to uncertainties in model inputs is extremely complex, and varies rapidly, perhaps more so than those in the policy arena would have previously thought. These new insights will aid in designing experiments to more thoroughly assess the effects of possible future emissions reduction policies.

Emulators have also made Monte Carlo uncertainty quantification viable for the first time without the inaccuracies inherent in the extrapolation of local sensitivity coefficients used by RFMs (reduced form models). There remains work to do in ensuring that such uncertainty quantification is accurate, but nonetheless the first steps have been made towards producing confidence intervals which can be used in epidemiological studies of the health effects of air pollution.

The emulators which have been built are among the largest described in the literature in terms of number of input variables and were surprisingly accurate given this high dimensionality, as demonstrated by the leave-one-out cross val-

idation. This was no doubt due to the large numbers of CMAQ training runs performed, 18 - 20 per input variable, which would not have been possible without automating the process of modifying the input data and launching the model runs. This automation has been the major technical achievement, particularly in on-the-fly source code modification and compilation of new model executables, and also in terms of controlling the flow of the huge amount of input and output data used in a large set of model runs.

The sections which follow briefly expand on some of the points made above and offer some suggestions for improvements to the work which has been carried out.

6.1 Gaussian Process Emulation

The majority of the results presented could not have been produced without the use of emulation; CMAQ is simply too computationally expensive to allow sufficient numbers of runs to be made. An exception is the bivariate plots of ozone response to changing NO_x and VOC emissions in section 4.4. These plots are actually surfaces constructed from 400 points each, so could have been made with 400 CMAQ runs, but were actually made with emulators trained on 576 runs. The advantage of emulation in this case comes when the analysis is changed; in order to swap NO_x emissions for NO emissions, for example, the same emulator could be used, but one would have to start the whole process from scratch if using CMAQ directly. As stated above, the model is among the more complex of those which have been emulated in the literature and the accuracy with which emulation has been performed was allowed by automating the process of performing many model runs. This in turn was made possible by the simple realization that compiling the model to run on a single processor and performing a number of runs at the same time would be considerably more efficient than compiling the model to run in parallel in the usual way and then performing the runs consecutively. Keeping track of the large volume of input and output data involved when running the

model in this way involved the careful construction of automated systems - the actual amount of data varied between different sets of runs, but typically a set of a few hundred ten day runs involves several terabytes of data contained in hundreds of thousands of separate files, most of which were created and deleted on the fly.

Apart from the results describing the effect of deposition velocity alone, it must be borne in mind that all of the analyses presented here are actually analyses of the emulators, and not of CMAQ itself. The differences between those results and the results which would be obtained if it were possible to carry out the same analyses on CMAQ directly have hopefully been minimised by producing emulators with a high degree of accuracy. As the project progressed and emulators with gradually higher numbers of input variables were produced, validation tests became more and more important, and some of these are described in chapter 4, showing that most of the emulators tested could predict CMAQ NO_2 output to within 1 ppb, and ozone to within 2 ppb. This process could have been improved upon, however, by including the calculation of validation metrics in automated parts of the analyses, and flagging up those emulators which failed any particular test.

The DiceKriging package was used throughout, with all parameters kept at default values, and emulators were simply constructed using large enough numbers of model runs to allow their accuracy to be validated. As the project progressed, and the author's knowledge of Gaussian process emulation improved, the package could have been used in a more intelligent way, and perhaps allowed emulators of similar accuracy to be constructed using fewer model runs. This was not done, however, in order to maintain consistent methodology throughout the project. A similar point can be made about LHS designs used to select the values of input variables for the emulator training runs. Several R packages exist which can produce some form of optimised LHS, commonly by producing a large number of designs and then selecting the best from among them according to some form

of optimality criterion. The project was started using a simple LHS to produce each emulator, and so was continued in the same fashion, again for the sake of continuity.

6.2 Input Variable Screening

A total of 223 input variables were screened using Morris' method in order to produce a short list of 31 which were carried forward for emulation. This was performed in two stages, as the capability to perturb reaction rates was not developed until part way through the project, but ideally would have been carried out in one process.

Using R3 and R28 (see table 3.3) to compare screening thresholds for reaction rates and emissions/BCs may not have the best thing to do as some photolysis rates, especially NO_2, turned out to be more significant. However, this does not appear to have adversely affected the analysis as all FASTs show some variables from both reaction rates and emissions with very small effects.

Given that the sensitivity analyses carried out after the screening process produced results with a high degree of spatio-temporal variability, where no single variable or even small group of variables dominates the uncertainty in the model output, it could be said that the screening was one of the most informative parts of the whole exercise. It is also worth noting that Morris' method can be used as a sensitivity analysis tool in its own right rather than just to rule variables in or out as it was used here. This is especially useful for models which have large numbers of input variables when a simple ranking of variables in order of importance may be easier to interpret than more complex methods, especially as the means of producing that ranking is easy for a broad audience to understand. Considering these facts it is surprising that Morris' method does not appear more frequently in the literature.

6.3 FAST

It is believed that the time series of main effects which have been produced for a number of locations represent the most comprehensive sensitivity analysis which has been performed with CMAQ to date. It is not possible to fully summarize the complex nature of the results in a few sentences, rather the figures themselves provide the best means of doing so. A few points worthy of note, however, are given here. The uncertainty in ozone BCs completely dominate the modelled ozone uncertainty during windy periods, but this influence dwindles to a comparatively negligible amount during the calmer conditions associated with pollution episodes. For the UK wide ozone concentration, emissions have very little influence, but this grows for rural areas in England and again for urban areas, with uncertainty in NO_x emissions dominating over uncertainty in VOC emissions, and biogenic VOCs being much more important than anthropogenic ones. The NO_2 sensitivity analyses were generally much less interesting, the modelled concentrations mainly being sensitive to UK NO emissions. There were, however, points in time where chemical reaction rates and ozone boundary conditions play a larger role, particularly in London, reflecting the complex nature of urban atmospheric chemistry.

The knowlege gained with such sensitivity analyses will assist the design of policy relevant experiments into the effects of emissions reductions. Rather than simply performing a few model runs with reduced NO_x or VOC emissions, the effect of reducing either of these at different times or in different geographical locations could be tested. Taking this further, it would be possible to perturb emissions by different amounts at different locations along the back trajectory from a given receptor location and thereby perform sensitivity analyses to determine whether air quality at that particular receptor was more sensitive to local emissions or long range transport, for example. A possible next step might be to group different times of arrival of air masses at particular locations by back trajectory and then perform the FAST on the aggregated model output for each

group of trajectories, perturbing those variables which were most important for a particular trajectory independently in separate groups of grid cells along that back trajectory, so allowing the sensitivity to uncertainty or sensitivity to emissions reductions at different points back in time from the location to be analysed.

The test of the sensitivity of FAST to the choice of input distributions showed that this can have an important affect on the results. If a more effective observational metric or set of metrics can be found with which to perform the history matching process this should be done before the FAST.

6.3.1 Elicitation of Expert Opinion

Related to the point above, a formal elicitation exercise on the distributions of the uncertain model inputs was not carried out during this project. Some of the studies mentioned in chapter 1 have attempted this, but all experienced difficulties of some kind or other. Hanna et al. (2001), for example, found that only around 20 % of the experts they contacted gave a response, and that those experts thought that the inputs they had been asked to consider must be correlated, but they could not specify how. A more formal framework, the Sheffield Elicitation Framework (SHELF), was used by Lee et al. (2013), but this by necessity involves far fewer experts, in this case six, who were able to give ranges for the parameters, but were unsure about specifying the shape of the distributions.

With hindsight, despite its difficulties, elicitation would be useful in adding some extra information, no matter how vague, especially in cases were no other source was available, for example on uncertainty surrounding deposition velocity. If a fully rigorous fully Bayesian calibration exercise became possible at some point in the future, the results of an elicitation exercise would help to formulate informative prior distributions for those model inputs which the observational data turned out not have enough power to shape.

It is worth mentioning, however, that part of an elicitation process, for models such as this with many inputs, is not just defining the shape of their distributions

but also identifying which variables should be included in the first place. This was, at least in part, replaced by the screening process.

6.4 Uncertainty in Meteorological Inputs

There should always be an attempt to minimise the errors in the meteorological inputs to a CTM with some kind of data assimilation technique, but the success of this inevitably depends on the quality of the data which is assimilated. The difference in spatial scale between the available data and the CMAQ grid size means small scale errors in wind fields may remain. The increased availability of higher resolution meteorological reanalysis data and advances in data assimilation hopefully means that these issues will become smaller over time.

The project progressed by gradually adding the capability to incorporate more input variables into the analysis, and naturally this started with those variables which were easiest to perturb. As one of the more difficult sets of inputs, meteorological variables may have been next on the list had the project continued for longer, and it may be that the same techniques used for data assimilation could provide the key to perturbing them without causing mass conservation issues. This may or may not be a desirable way to proceed, however, as meteorological uncertainty would be useful to include in sensitivity analysis, but in calibration would only serve to confound the equifinality problem, and here it may be better to focus instead on improving the meteorological data as much as possible.

6.5 History Matching and Calibration

As alluded to in the previous chapter the first thing which must be done before attempting any kind of serious calibration exercise is to identify the correct observational metric against which to compare the model output. It would seem logical that the most likely way to alleviate the problems caused by equifinality is to use observations of several species at once. In particular to look at a time

period when measurements of ozone, NO, NO_2 and isoprene were all available would help to calibrate modelled ozone and NO_2 because of their highly intertwined chemistry. The Harwell monitoring site is supposed to measure all of these species continuously, but in some instances there is a lot of missing data. The isoprene measurement record from 1/1/2006 to 31/12/2010, for example, contains 85 % missing data, but nonetheless the time spent searching the record for a suitable period may prove worthwhile.

There are still unaccounted for uncertainties which may hinder a calibration process, such as the remaining uncertainties in meteorological data which will exist even after the best attempts at data assimilation have been made. If these uncertainties can be assumed to be random then using aggregated observational metrics, such as means over time and/or space might negate their effects. One disadvantage of using average values, however, is that they mask natural variability, such as the diurnal cycle of ozone concentration which is a fundamental feature of the model output. It may be then that different methods of summarizing large numbers of observations are more appropriate. Perhaps through principal components analysis of a number of time series of observations at different monitoring sites and then emulating the principle components. A similar approach was taken in the climate model calibration by Sexton et al. (2011), who performed a singular value decomposition of 175,000 observations and emulated the eigenvectors that this produced. Whichever approach is taken to summarize the observational data, a successful calibration is much more likely if it is performed as an experiment designed specifically for that purpose, rather than the approach taken here, which was more of an addition to the existing sensitivity analysis.

In spite of the points made above, the rather crude calibration exercise which was performed did in fact produce significant improvements, particularly in predictions of afternoon peak ozone concentrations. This should be extended to produce separate sets of calibration factors for daytime and night-time, and these

should be tested with the CMAQ model itself rather than just with the emulators. The calibration factors, or 'scaling factors', which were calculated during the calibration process can provide useful information in themselves, even though it has been acknowledged that they may not reflect the real error in a particular factor, if that input factor has been over or under calibrated to make up for some unaccounted-for uncertainty. However, they can provide supporting evidence if it is already suspected that a model input factor might be incorrect. For example, the scaling factors in table 5.1 suggest that NO emissions should be 30 - 40 % higher and that NO_2 emissions should be 25 - 30 % higher. This is in broad agreement with on-road measurement campaigns of motor vehicle exhaust, for example Carslaw and Rhys-Tyler (2013), which suggest that real world NO_x emissions may be significantly higher than official figures.

6.6 In Summary

With careful automation it is possible to peform enough runs to emulate the CMAQ model and thereby create quick-to-run model surrogates which are significantly more accurate than simpler reduced form models. Central to this is the understanding that simultaneous single-processor runs will always be faster than consecutive many-processor runs.

Morris' method is highly effective for screening large numbers of input variables and should be recommended as a first step for anyone approaching the analysis of a model with many inputs.

Global sensitivity analysis performed at many locations and time points reveals that the input uncertainties which have the most influence over output uncertainty change rapidly with place and time. Some major points to reiterate are that ozone boundary conditions are the dominant influence on modelled ozone concentrations during windy conditions but in the calmer periods associated with pollution episodes NO_x emissions have considerably more influence. Also, uncertainty in biogenic VOC emissions appears to be far more influential

than uncertainty in anthropogenic VOC emissions. Care should be taken when making policy recommendations based on the results of the FAST analyses as they are primarily tests of sensitivity to input uncertainty. However the bivariate ozone isopleths in the same chapter and in appendix A add support to the argument that biogenic VOC emmisions are more important than anthropogenic ones, and so NO_x emission controls may be more helpful in curbing ozone pollution episodes. Incorporating uncertainty in meteorological inputs into sensitivity analyses should be one of the key aims of any continuation of the work presented here, and in particular in investigation into whether data assimilation methods may help in this regard should be carried out.

It has been demonstrated that history matching and calibration procedures for the CMAQ model can be made possible with the use of Gaussian process emulation, where before this kind of operation could not even be considered. The next stage in this research should be the search for the most appropriate set of observational metrics with which to perform further calibration exercises.

Bibliography

AEA (2007). QA/QC Data Ratification and Intercalibration Report for the Automatic Urban and Rural Network, July-September 2006. Technical report, AEA energy and environment, Didcot.

Anderson, H., Atkinson, R., Bremner, S., Carrington, J., and Peacock, J. (2007). Quantitative systematic review of short term associations between ambient air pollution (particulate matter, ozone, nitrogen dioxide, sulphur dioxide and carbon monoxide), and mortality and morbidity. Technical report, Department of Health, London.

Anderson, T. (1962). On the distribution of the two-sample Cramér-von Mises criterion. *The Annals of Mathematical Statistics*, 33(3):1148–1159.

Andersson, C. and Engardt, M. (2010). European ozone in a future climate: Importance of changes in dry deposition and isoprene emissions. *Journal of Geophysical Research*, 115(D2):D02303.

Appel, K. W., Gilliland, A. B., Sarwar, G., and Gilliam, R. C. (2007). Evaluation of the Community Multiscale Air Quality (CMAQ) model version 4.5: Sensitivities impacting model performance Part I - Ozone. *Atmospheric Environment*, 41(40):9603–9615.

Appel, K. W., Roselle, S. J., Gilliam, R. C., and Pleim, J. E. (2010). Sensitivity of the Community Multiscale Air Quality (CMAQ) model v4.7 results for the eastern United States to MM5 and WRF meteorological drivers. *Geoscientific Model Development*, 3(1):169–188.

Arnold, J. R. and Dennis, R. L. (2006). Testing CMAQ chemistry sensitivities in base case and emissions control runs at SEARCH and SOS99 surface sites in the southeastern US. *Atmospheric Environment*, 40(26):5027–5040.

Atkinson, R., Anderson, H., Sunyer, J., Ayres, J., Baccini, M., Vonk, J., Boumghar, A., Forastiere, F., Forsberg, B., Touloumi, G., Schwartz, J., and Katsouyanni, K. (2001). Acute effects of particulate air pollution on respiratory

admissions: results from APHEA 2 project. *American journal of respiratory and critical care medicine*, 164:1860–1866.

Bastos, L. S. and O'Hagan, A. (2009). Diagnostics for Gaussian Process Emulators. *Technometrics*, 51(4):425–438.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. a., Cavendish, J., Lin, C.-H., and Tu, J. (2007). A Framework for Validation of Computer Models. *Technometrics*, 49(2):138–154.

Beekmann, M. and Derognat, C. (2003). Monte Carlo uncertainty analysis of a regional-scale transport chemistry model constrained by measurements from the Atmospheric Pollution Over the Paris Area (ESQUIF) campaign. *Journal of Geophysical Research*, 108(D17).

Bergin, M. S. and Milford, J. B. (2000). Application of Bayesian Monte Carlo analysis to a Lagrangian photochemical air quality model. *Atmospheric Environment*, 34(5):781–792.

Bergin, M. S., Noblet, G. S., Petrini, K., Dhieux, J. R., Milford, J. B., and Harley, R. a. (1999). Formal Uncertainty Analysis of a Lagrangian Photochemical Air Pollution Model. *Environmental Science & Technology*, 33(7):1116–1126.

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010). A Spatio-Temporal Downscaler for Output From Numerical Models. *Journal of agricultural, biological, and environmental statistics*, 15(2):176–197.

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1-2):18–36.

Beven, K. and Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological processes*, 6:279–298.

Byun, D. W. (1999a). consistent formulations in meteorological and air quality models for multiscale atmospheric studies. Part I: Governing equations in a generalized coordinate system. *Journal of the atmospheric sciences*, (Byun):3789–3807.

Byun, D. W. (1999b). Dynamically consistent formulations in meteorological and air quality models for multiscale atmospheric studies. Part II: Mass conservation issues. *Journal of the Atmospheric Sciences*, pages 3808–3820.

Byun, D. W. and Schere, K. L. (2006). Review of the governing equations, computational algorithms, and other components of the models-3 Community

Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews*, 59(1-6):51–77.

Campolongo, F., Cariboni, J., and Saltelli, A. (2007). An effective screening design for sensitivity analysis of large models. *Environmental Modelling & Software*, 22(10):1509–1518.

CARB (2007). Review of the California Ambient Air Quality Standard for Nitrogen Dioxide. Technical report, California Environmental Protection Agency Air Resources Board, Sacramento.

Carslaw, D., Agnew, P., Beevers, S., Chemel, C., Cooke, S., Davis, L., Derwent, D., Francis, X., Fraser, A., Kitwiroon, N., McFiggans, G., Murrels, T., Redington, A., Savage, N., Sokhi, R., and Vieno, M. (2013). DEFRA Phase 2 Regional Model Evaluation. Technical report, DEFRA.

Carslaw, D. C. and Rhys-Tyler, G. (2013). New insights from comprehensive on-road measurements of NOx, NO2 and NH3 from vehicle emission remote sensing in London, UK. *Atmospheric Environment*, 81(2):339–347.

CEDA (2015). Centre for environmental data archival. [Online] available at: http://ceda.ac.uk [accessed 15/04/2015].

CMAS (2010). Operational guidance for the Community Multiscale Air Quality (CMAQ) modeling system. Technical Report June, University of North Carolina, Chapel Hill.

CMAS (2014). The community modelling and analysis system. [Online] available at: https://www.cmascenter.org/index.cfm [accessed 20/01/2015].

Cohan, D. S., Hakami, A., Hu, Y., and Russell, A. G. (2005). Nonlinear response of ozone to emissions: source apportionment and sensitivity analysis. *Environmental science & technology*, 39(17):6739–48.

Cohan, D. S., Tian, D., Hu, Y., and Russell, A. G. (2006). Control strategy optimization for attainment and exposure mitigation: case study for ozone in Macon, Georgia. *Environmental management*, 38(3):451–62.

Conti, S. and O'Hagan, A. (2010). Bayesian emulation of complex multi-output and dynamic computer models. *Journal of Statistical Planning and Inference*, 140(3):640–651.

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. H. (1996). Bayes linear strategies for matching hydrocarbon reservoir history. In Bernardo, J. M.,

Berger, J. O., Dawid, A. P., and Smith, A. F. M., editors, *Bayesian statistics 5*, pages 69–95. Oxford University Press, Oxford.

Cressie, N. (1993). *statistics for spatial data*. Wiley.

Cukier, R. I., Levine, H. B., and Shuler, K. E. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26:1–42.

DEFRA (2014). Uk-air data archive. [Online] available at: http://uk-air.defra.gov.uk/data/ [accessed 20/01/2015].

Deguillaume, L., Beekmann, M., and Derognat, C. (2008). Uncertainty evaluation of ozone production and its sensitivity to emission changes over the Ile-de-France region during summer periods. *Journal of Geophysical Research*, 113(D2):D02304.

Deguillaume, L., Beekmann, M., and Menut, L. (2007). Bayesian Monte Carlo analysis applied to regional-scale inverse emission modeling for reactive trace gases. *Journal of Geophysical Research*, 112(D2):D02307.

Derwent, D., Fraser, A., Abbott, J., Jenkin, M., Willis, P., and Murrells, T. (2010). Evaluating the Performance of Air Quality Models. Technical Report 3, DEFRA.

Derwent, R., Jenkin, M., and Saunders, S. (1996). Photochemical ozone creation potentials for a large number of reactive hydrocarbons under European conditions. *Atmospheric Environment*, 30(2):181–199.

Derwent, R. G. and Murrells, T. P. (2013). Impact of policy-relevant scenarios on ozone in southern England: Influence of chemical mechanism choice. *Atmospheric Environment*, 72:89–96.

Digar, A., Cohan, D. S., Xiao, X., Foley, K. M., Koo, B., and Yarwood, G. (2013). Constraining ozone-precursor responsiveness using ambient measurements. *Journal of Geophysical Research: Atmospheres*, 118(2):1005–1019.

Dilks, D., Canale, R., and Meier, P. (1992). Development of Bayesian Monte Carlo techniques for water quality model uncertainty. *Ecological Modelling*, 62:149–162.

Dunker, A. M. (1984). The decoupled direct method for calculating sensitivity coefficients in chemical kinetics. *The Journal of Chemical Physics*, 81(5):2385–2393.

Dupuy, D. and Helbert, C. (2013). *DiceEval: Construction and evaluation of metamodels*. R package version 1.2.

Edwards, N. R., Cameron, D., and Rougier, J. (2010). Precalibrating an intermediate complexity climate model. *Climate Dynamics*, 37(7-8):1469–1482.

Emberson, L. D., Kitwiroon, N., Beevers, S., Büker, P., and Cinderby, S. (2013). Scorched Earth: how will changes in the strength of the vegetation sink to ozone deposition affect human health and ecosystems? *Atmospheric Chemistry and Physics*, 13(14):6741–6755.

ENVIRON (2010). User's guide-comprehensive air-quality model with extensions, Version 5.30. Technical report, ENVIRON International Corporation, Novato, California.

EU (2001). 2001/81/EC on national emissions ceilings for certain atmospheric pollutants. *Official Journal of the European Union*, L309/22.

EU (2008). 2008/50/EC on ambient air quality and cleaner air for Europe. *Official Journal of the European Union*, L153/1.

Fang, K., Li, R., and Sudjianto, A. (2006). *Design and modeling for computer experiments*. Chapman & Hall/CRC, Boca Raton.

Faraji, M., Kimura, Y., Allen, D., McDonald-Buller, E., and Allen, D. (2008). Comparison of the carbon bond and SAPRC photochemical mechanisms under conditions relevant to southeast Texas. *Atmospheric Environment*, 42(23):5821–5836.

Ferguson, N. and Harrison, R. (2010). Science Advisory Council: Modelling, Uncertainty and Prediction. Technical Report October, DEFRA.

Finlayson-Pitts, B. and Pitts, J. (2000). *Chemistry of the Upper and Lower Atmosphere*. Academic Press, San Diego.

Gonçalves, M., Dabdub, D., Chang, W., Jorba, O., and Baldasano, J. (2012). Impact of HONO sources on the performance of mesoscale air quality models. *Atmospheric Environment*, 54(2):168–176.

Hahmann, A. N., Rostkier-Edelstein, D., Warner, T. T., Vandenberghe, F., Liu, Y., Babarsky, R., and Swerdlin, S. P. (2010). A reanalysis system for the generation of mesoscale climatographies. *Journal of Applied Meteorology and Climatology*, 49(5):954–972.

Hakami, A., Odman, M. T., and Russell, A. G. (2003). High-order, direct sensitivity analysis of multidimensional air quality models. *Environmental science & technology*, 37(11):2442–52.

Hankin, R. K. S. (2005). Introducing BACCO, an R Bundle for Bayesian Analysis of Computer Code Output. *journal of statistical software*, 14(16).

Hanna, S. R., Lu, Z., Frey, H. C., Wheeler, N., Vukovich, J., Arunachalam, S., Fernau, M., and Hansen, D. A. (2001). Uncertainties in predicted ozone concentrations due to input uncertainties for the UAM-V photochemical grid model applied to the July 1995 OTAG domain. *Atmospheric Environment*, 35(July 1995):891–903.

Isukapalli, S. S., Roy, a., and Georgopoulos, P. G. (1998). Stochastic response surface methods (SRSMs) for uncertainty propagation: application to environmental and biological systems. *Risk analysis : an official publication of the Society for Risk Analysis*, 18(3):351–63.

Isukapalli, S. S., Wang, S. W., Lahoti, N., Unal, A., and Georgopoulos, P. G. (2005). Efficient Techniques for Sensitivity and Uncertainty Analysis of Multiscale Air Quality Models. In *4th Annual CMAS Models-3 User's Conference, UNC-Chapel Hill*.

Jacob, D. J. (1999). *Introduction to Atmospheric Chemistry*. Princeton University Press, Princeton.

Jacobson, M. (2002). *Atmospheric Pollution: History, Science and Regulation*. Cambridge University Press, Cambridge.

Jenkin, M. E. and Clemitshaw, K. C. (2000). Ozone and other secondary photochemical pollutants: chemical processes governing their formation in the planetary boundary layer. *Atmospheric Environment*, 34(16):2499–2527.

Jerrett, M., Burnett, R., Pope, C., Ito, K., Thurston, G., Krewski, D., Shi, Y., Calle, E., and Thun, M. (2009). Long-term ozone exposure and mortality. *New England journal of Medicine*, 360:1085–1095.

Ji, M., Cohan, D. S., and Bell, M. L. (2011). Meta-analysis of the Association between Short-Term Exposure to Ambient Ozone and Respiratory Hospital Admissions. *Environmental Research Letters*, 6(2).

Jiménez, P., Parra, R., and Baldasano, J. M. (2007). Influence of initial and boundary conditions for ozone modeling in very complex terrains: A case study

in the northeastern Iberian Peninsula. *Environmental Modelling & Software*, 22(9):1294–1306.

Jin, L., Tonse, S., Cohan, D., Mao, X., Harley, R. a., and Brown, N. (2008). Sensitivity analysis of ozone formation and transport for a central California air pollution episode. *Environmental Science & Technology*, 42:3683–3689.

Katsouyanni, K., Samet, J., Anderson, H., Atkinson, R., Le Tertre, A., Medina, S., Samoli, E., Touloumi, G., Burnett, R., Krewski, D., Ramsay, T., Dominici, F., Peng, R., Schwartz, J., and Zanobetti, A. (2009). Air Pollution and Health: A European and North American Approach (APHENA). HEI Research Report 142. Technical report, Health Effects Institute, Boston, MA.

Kennedy, M. C., Anderson, C. W., Conti, S., and O'Hagan, A. (2006). Case studies in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety*, 91(10-11):1301–1309.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

Laffineur, Q., Aubinet, M., Schoon, N., Amelynck, C., Müller, J.-F., Dewulf, J., Van Langenhove, H., Steppe, K., Šimpraga, M., and Heinesch, B. (2011). Isoprene and monoterpene emissions from a mixed temperate forest. *Atmospheric Environment*, 45(18):3157–3168.

Lee, J., Lewis, a., Monks, P., Jacob, M., Hamilton, J., Hopkins, J., Watson, N., Saxton, J., Ennis, C., and Carpenter, L. (2006). Ozone photochemistry and elevated isoprene during the UK heatwave of august 2003. *Atmospheric Environment*, 40(39):7598–7613.

Lee, L. A., Carslaw, K. S., Pringle, K. J., and Mann, G. W. (2012). Mapping the uncertainty in global CCN using emulation. *Atmospheric Chemistry and Physics*, 12(20):9739–9751.

Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., Pierce, J. R., and Carslaw, K. S. (2013). The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmospheric Chemistry and Physics*, 13(17):8879–8914.

Li, G., Rabitz, H., Hu, J., Chen, Z., and Ju, Y. (2008). Regularized random-sampling high dimensional model representation (RS-HDMR). *Journal of Mathematical Chemistry*, 43(3):1207–1232.

Li, G. Y., Rosenthal, C., and Rabitz, H. (2001). High dimensional model representations. *Journal of Physical Chemistry A*, 105(33):7765–7777.

Luecken, D., Phillips, S., Sarwar, G., and Jang, C. (2008). Effects of using the CB05 vs. SAPRC99 vs. CB4 chemical mechanism on model predictions: Ozone and gas-phase photochemical precursor concentrations. *Atmospheric Environment*, 42(23):5805–5820.

Madden, M. C. and Hogsett, W. E. (2001). A Historical Overview of the Ozone Exposure Problem. *Human and Ecological Risk Assessment: An International Journal*, 7(5):1121–1131.

Mathur, R. (2005). Multiscale Air Quality Simulation Platform (MAQSIP): Initial applications and performance for tropospheric ozone and particulate matter. *Journal of Geophysical Research*, 110(D13):D13308.

McKay, M., Beckman, R., and Conover, W. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.

McNair, L., Harley, R., and Russell, A. (1996). Spatial inhomogeneity in pollutant concentrations, and their implications for air quality model evaluation. *Atmospheric Environment*, 30(24):4291–4301.

McNeall, D. J., Challenor, P. G., Gattiker, J. R., and Stone, E. J. (2013). The potential of an observational data set for calibration of a computationally expensive computer model. *Geoscientific Model Development*, 6(5):1715–1728.

Morris, M. D. (1991). Factorial Sampling Plans for Preliminary Computational Experiments. *Technometrics*, 33(2):161–174.

Mueller, S. F. and Mallard, J. W. (2011). Contributions of Natural Emissions to Ozone and PM(2.5) as Simulated by the Community Multiscale Air Quality (CMAQ) Model (vol 45, pg 4817, 2011). *Environmental Science & Technology*, 45(18):7950.

Napelenok, S., Cohan, D., Hu, Y., and Russell, a. (2006). Decoupled direct 3D sensitivity analysis for particulate matter (DDM-3D/PM). *Atmospheric Environment*, 40(32):6112–6121.

Napelenok, S., Cohan, D., Odman, M., and Tonse, S. (2008). Extension and evaluation of sensitivity analysis capabilities in a photochemical model. *Environmental Modelling & Software*, 23(8):994–999.

NOAA (2015). National oceanic and atmospheric administration global forecast system. [Online] available at: http://ncdc.noaa.gov/data-access/model-data/model-datasets/global-forcast-system-gfs [accessed 17/04/2015].

NOAA Air Resources Laboratory (2015). Hysplit - hybrid single particle lagrangian integrated trajectory model. [Online] available at: http://ready.arl.noaa.gov/HYSPLIT.php [accessed 15/04/2015].

Oakley, J. and O'Hagan, A. (2002). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, 89(4):769–784.

Oakley, J. E. and O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769.

O'Hagan, A. (2006). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety*, 91(10-11):1290–1300.

Otte, T. L. and Pleim, J. E. (2010). The Meteorology-Chemistry Interface Processor (MCIP) for the CMAQ modeling system: updates through MCIPv3.4.1. *Geoscientific Model Development*, 3(1):243–256.

Passant, N. (2003). Estimation of Uncertainties in the National Atmospheric Emissions Inventory. Technical report, AEA Technology, Abingdon.

Passant, N., Murrells, T., Pang, Y., Thistlethwaite, G., Walker, C., Watterson, J., Webb, N., Broomfield, M., Mitchell, R., Webb, J., Hunter, R., Bailey, R., Buckland, T., Hobson, M., Dore C, and Misselbrook, T. (2014). UK Informative inventory Report (1980 to 2012): Annual Report for Submission Under the UNECE-Convention on Long Range Trans Boundary Air Pollution. Technical report, DEFRA, London.

Pinder, R. W., Gilliam, R. C., Appel, K. W., Napelenok, S. L., Foley, K. M., and Gilliland, A. B. (2009). Efficient Probabilistic Estimates of Surface Ozone Concentration Using an Ensemble of Model Configurations and Direct Sensitivity Calculations. 43(7):2388–2393.

Pleim, J. and Ran, L. (2011). Surface Flux Modeling for Air Quality Applications. *Atmosphere*, 2(4):271–302.

Pujol, G., Iooss, B., with contributions from Laurent Gilquin, A. J., Gratiet, L. L., and Lemaitre, P. (2014). *sensitivity: Sensitivity Analysis*. R package version 1.8-1.

Pukelsheim, F. (1994). The Three Sigma Rule. *the american statistician*, 48(2):88–91.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rao, S. T., Galmarini, S., and Puckett, K. (2011). Air Quality Model Evaluation International Initiative (AQMEII): Advancing the State of the Science in Regional Photochemical Modeling and Its Applications. *Bulletin of the American Meteorological Society*, 92(1):23–30.

Rasmussen, C. E. and Williams, K. I. (2006). *Gaussian Processes for Machine Learning*. MIT Press, Cambridge.

REVIHAAP (2013). Review of evidence on health aspects of air pollution. Technical report, WHO, Copenhagen.

Rojas-Martinez, R., Perez-Padilla, R., Olaiz-Fernandez, G., Mendoza-Alvarado, L., Moreno-Macias, H., Fortoul, T., McDonnell, W., Loomis, D., and Romieu, I. (2007). Lung function growth in children with long-term exposure to air pollutants in Mexico City. *American journal of respiratory and critical care medicine*, 176(4):377–84.

Rougier, J. (2007). Probabilistic Inference for Future Climate Using an Ensemble of Climate Model Evaluations. *Climatic Change*, 81:247–264.

Roustant, O., Ginsbourger, D., and Deville, Y. (2012). DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. *Journal of Statistical Software*, 51(1):1–55.

Sacks, J., Welch, W., Mitchell, T., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical science*, 4(4):409–423.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270.

Saltelli, A., Chan, K., and Scott, E. (2008). *Sensitivity Analysis*. Wiley, Chichester.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice*. Wiley, Chichester.

Saltelli, A., Tarantola, S., and Chad, K. (1998). Presenting Results from Model Based Studies to Decision-Makers: Can Sensitivity Analysis Be a Defogging Agent? *Risk Analysis*, 18(6):799–803.

Saltelli, A., Tarantola, S., and Chan, K. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1):39–56.

Sander, S., Abbatt, J., Barker, J., Burkholder, J., Friedl, R., Golden, D., Huie, R., Kolb, C., Kurylo, M., Moortgat, G., Orkin, V., and Wine, P. (2011). Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation No.17, JPL Publication 10-6. Technical report, Jet Propulsion Laboratory, Pasadena.

Santner, T., Williams, B., and Notz, W. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York.

Sarwar, G., Luecken, D., Yarwood, G., Whitten, G. Z., and Carter, W. P. L. (2008). Impact of an updated carbon bond mechanism on predictions from the CMAQ modeling system: Preliminary assessment. *Journal of Applied Meteorology and Climatology*, 47(1):3–14.

Sexton, D. M. H. and Murphy, J. M. (2011). Multivariate probabilistic projections using imperfect climate models. Part II: robustness of methodological choices and consequences for climate sensitivity. *Climate Dynamics*, 38(11-12):2543–2558.

Sexton, D. M. H., Murphy, J. M., Collins, M., and Webb, M. J. (2011). Multivariate probabilistic projections using imperfect climate models part I: outline of methodology. *Climate Dynamics*, 38:2513–2542.

Sillman, S. (1999). The relation between ozone, NOx and hydrocarbons in urban and polluted rural environments. *Atmospheric Environment*, 33(12):1821–1845.

Sillman, S. and He, D. (2002). Some theoretical results concerning O3-NOx-VOC chemistry and NOx-VOC indicators. *Journal of Geophysical Research*, 107(D22):4659.

Sivia, D. S. and Skilling, J. (2006). *Data Analysis, a Bayesian Tutorial*. Oxford University Press, Oxford, 2 edition.

Smith, K. R., Jerrett, M., Anderson, H. R., Burnett, R. T., Stone, V., Derwent, R., Atkinson, R. W., Cohen, A., Shonkoff, S. B., Krewski, D., Pope, C. A., Thun, M. J., and Thurston, G. (2009). Public health benefits of strategies to

reduce greenhouse-gas emissions: health implications of short-lived greenhouse pollutants. *Lancet*, 374:2091–103.

Sobol', I. (1990). On sensitivity estimation for nonlinear mathematical models. *matematicheskoe modelirovanie*, 2(1):112–118.

Stedinger, J. R., Vogel, R. M., Lee, S. U., and Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resources Research*, 44(12):n/a–n/a.

Stedman, J. R. (2004). The predicted number of air pollution related deaths in the UK during the August 2003 heatwave. *Atmospheric Environment*, 38(8):1087–1090.

Stockwell, W., Middleton, P., Chang, J., and Tang, X. (1990). The second generation regional acid deposition model chemical mechanism for regional air quality modeling. *Journal of Geophysical Research*, 95:16343– 16376.

Strong, J., Whyatt, J. D., Metcalfe, S. E., Derwent, R. G., and Hewitt, C. N. (2013). Investigating the impacts of anthropogenic and biogenic VOC emissions and elevated temperatures during the 2003 ozone episode in the UK. *Atmospheric Environment*, 74(August 2003):393–401.

Tang, Y. H., Lee, P., Tsidulko, M., Huang, H. C., McQueen, J. T., DiMego, G. J., Emmons, L. K., Pierce, R. B., Thompson, A. M., Lin, H. M., Kang, D. W., Tong, D., Yu, S. C., Mathur, R., Pleim, J. E., Otte, T. L., Pouliot, G., Young, J. O., Schere, K. L., Davidson, P. M., and Stajner, I. (2009). The impact of chemical lateral boundary conditions on CMAQ predictions of tropospheric ozone over the continental United States. *Environmental Fluid Mechanics*, 9(1):43–58.

Tian, D., Cohan, D. S., Napelenok, S., Bergin, M., Hu, Y., Chang, M., and Russell, A. G. (2010). Uncertainty Analysis of Ozone Formation and Response to Emission Controls Using Higher-Order Sensitivities. *Journal of the Air & Waste Management Association*, 60(7):797–804.

Tzivian, L. (2011). Outdoor air pollution and asthma in children. *Journal of Asthma*, 48(5):470–81.

USEPA (2008). Integrated Science Assessment for Oxides of Nitrogen - Health Criteria. Technical report, United States Environmental Protection agency, Washington, DC.

USEPA (2011). U.S. EPA. Integrated Science Assessment of Ozone and Related Photochemical Oxidants (Second External Review Draft). Technical report, United States Environmental Protection Agency, Washington, DC.

Vernon, I., Goldstein, M., and Bower, R. G. (2010). Galaxy Formation: a Bayesian Uncertainty Analysis. *Bayesian Analysis*, 5(4):619–670.

Vuilleumier, L., Harley, R. a., and Brown, N. J. (1997). First- and Second-Order Sensitivity Analysis of a Photochemically Reactive System (a Green's Function Approach). *Environmental Science & Technology*, 31(4):1206–1217.

Wang, S. W., Georgopoulos, P. G., Li, G., and Rabitz, H. (2001). Computationally efficient atmospheric chemical kinetic modeling by means of high dimensional model representation (HDMR). In Margenov, S. Y. P. W. J., editor, *Large-Scale Scientific Computing*, volume 2179, pages 326–333. Springer-Verlag, Berlin.

Wang, S. W., Georgopoulos, P. G., Li, G. Y., and Rabitz, H. (2003). Random sampling-high dimensional model representation (RS-HDMR) with nonuniformly distributed variables: application to an integrated multimedia/multipathway exposure and dose model for trichloroethylene. *The Journal of Physical Chemistry A*, 107(23):4707–4716.

Whitten, G. Z., Hogo, H., and Killus, J. P. (1980). The carbon-bond mechanism: A condensed kinetic mechanism for photochemical smog. *Environmental Science & Technology*, 14(6):690–700.

Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K. (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate Dynamics*, 41(7-8):1703–1729.

Yang, Y.-J., Wilkinson, J. G., and Russell, A. G. (1997). Fast, Direct Sensitivity Analysis of Multidimensional Photochemical Models. *Environmental Science & Technology*, 31(10):2859–2868.

Yarwood, G., Rao, S., Yocke, M., and Whitten, G. (2005). UPDATES TO THE CARBON BOND CHEMICAL MECHANISM: CB05. Technical report, USEPA.

Yu, Y., Sokhi, R. S., Kitwiroon, N., Middleton, D. R., and Fisher, B. (2008). Performance characteristics of MM5-SMOKE-CMAQ for a summer photochemical episode in southeast England, United Kingdom. *Atmospheric Environment*, 42(20):4870–4883.

Zanobetti, A. and Schwartz, J. (2011). Ozone and survival in four cohorts with potentially predisposing diseases. *American journal of respiratory and critical care medicine*, 184(7):836–41.

Zhou, W., Cohan, D. S., and Napelenok, S. L. (2013). Reconciling NOx emissions reductions and ozone trends in the U.S., 2002–2006. *Atmospheric Environment*, 70(x):236–244.

Appendix A

Additional Results

The locations for which results are provided are all model grid squares which contain an air quality monitoring station. The descriptive information for each location is taken from the UK-Air website (http://uk-air.defra.gov.uk/).

A.1 Market Harborough

The monitoring station is located within a self contained, air conditioned housing. The nearest road is a rural road approximately 160 metres from the station. The surrounding area is rural arable farmland.

Environment Type: Rural Background

Altitude: 145 m

Latitude/Longitude: 52.554444, -0.772222

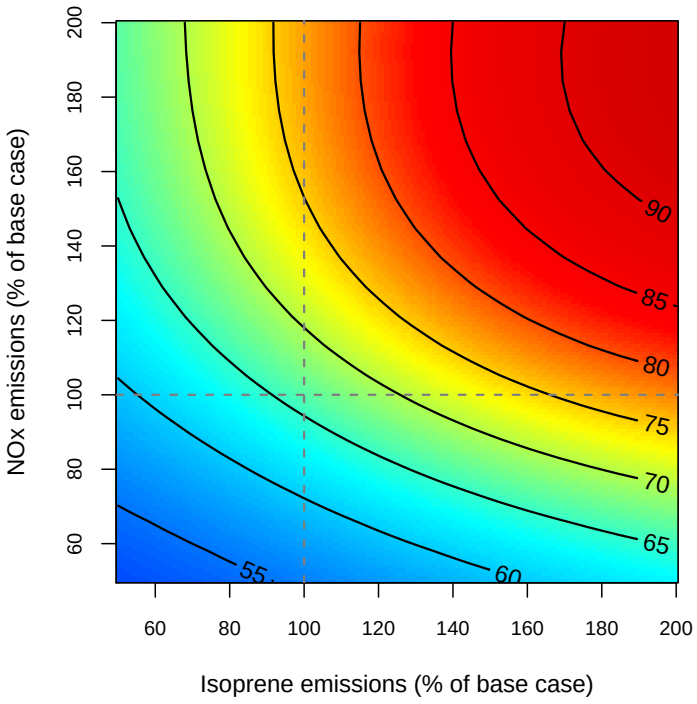


Figure A.1: Ozone concentrations at Market Harborough, labelled in ppb on the contours, against NO_x and isoprene emissions.

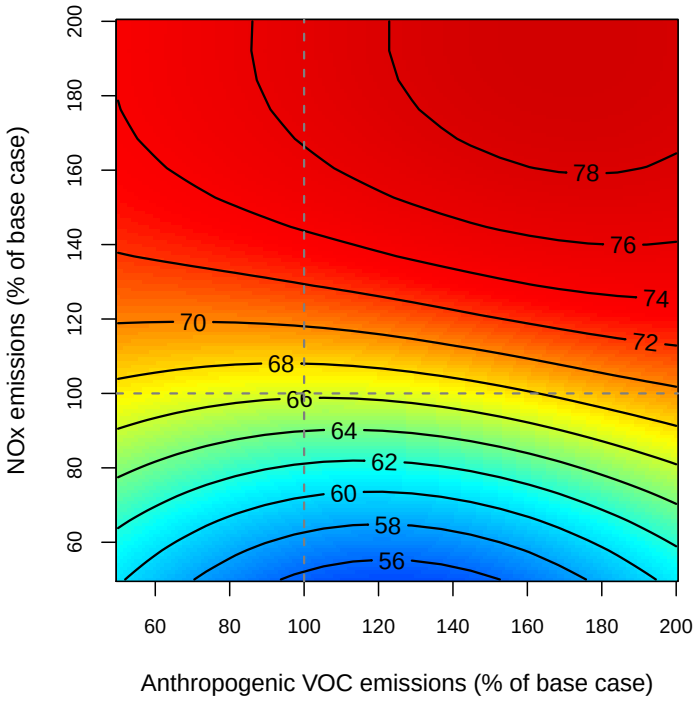


Figure A.2: Ozone concentrations at Market Harborough, labelled in ppb on the contours, against NO_x and anthropogenic VOC emissions.

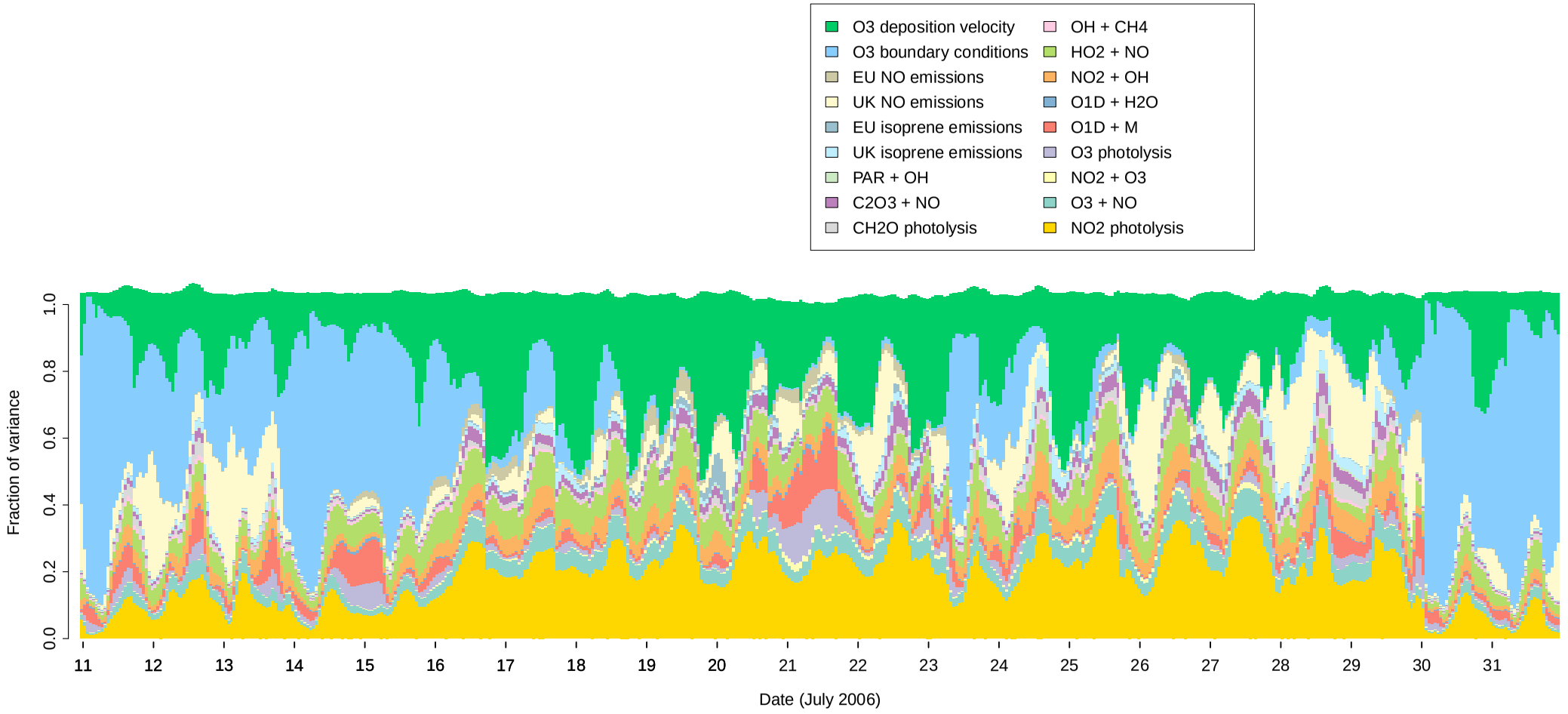


Figure A.3: Time series of Main effects on modelled ozone concentrations at Market Harborough, 11th-31st July

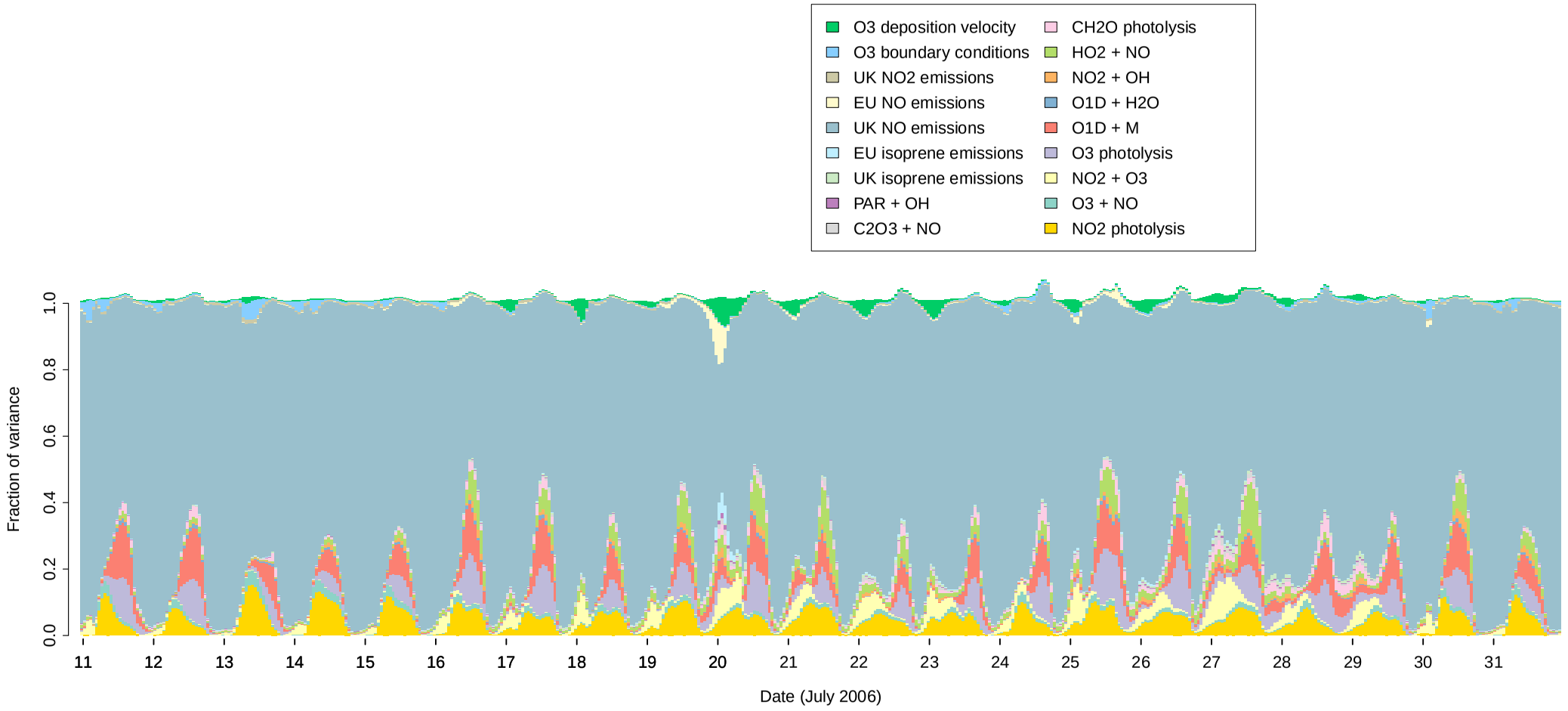


Figure A.4: Time series of Main effects on modelled NO_2 concentrations at Market Harborough, 11th-31st July

A.2 Ladybower

The monitoring station is located within a self-contained, air conditioned housing located in the Peak District National Park approximately 800 metres to the south west of Ladybower reservoir. The nearest road is for access to the nearby farm buildings only, and is approximately 20 metres from the station. The surrounding area is mainly open moor land with the nearest trees occurring within a distance of several hundred metres.

Environment Type: Rural Background

Altitude: 420 m

Latitude/Longitude: 53.403370, -1.752006

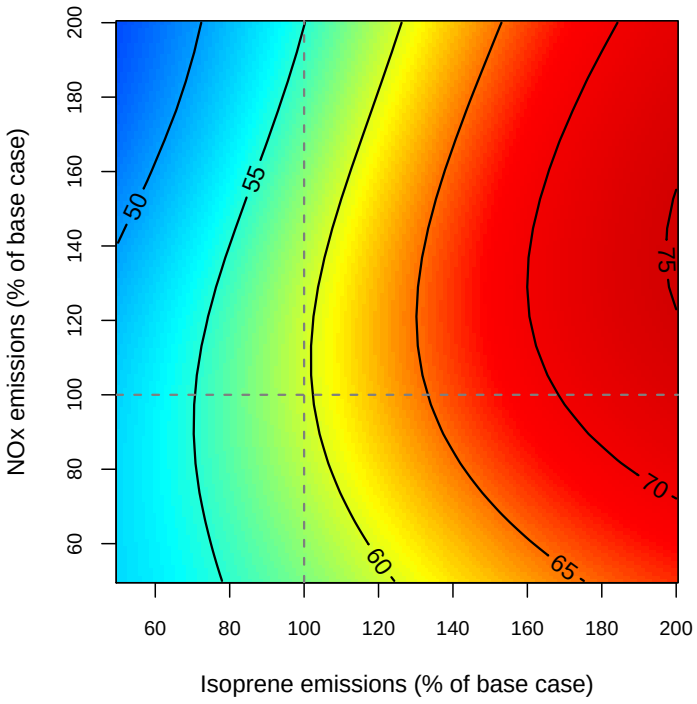


Figure A.5: Ozone concentrations at Ladybower, labelled in ppb on the contours, against NO_x and isoprene emissions.

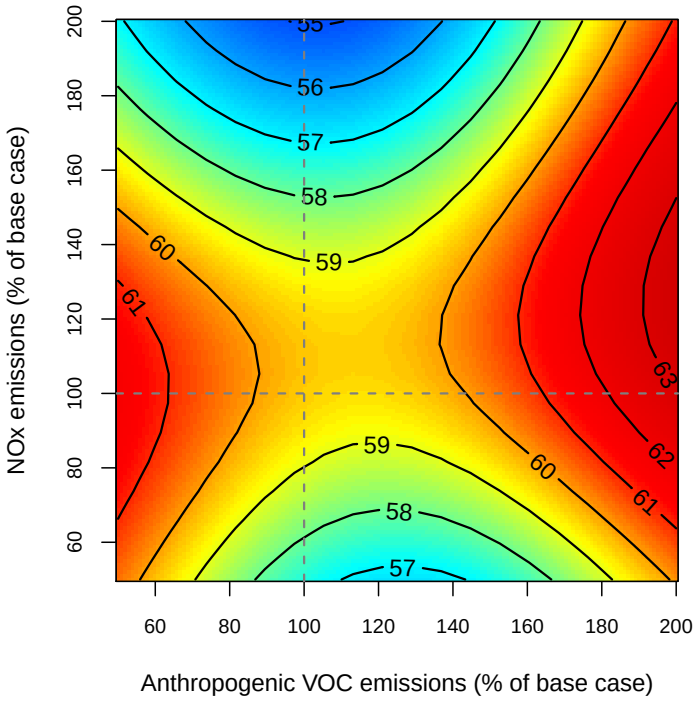


Figure A.6: Ozone concentrations at Ladybower, labelled in ppb on the contours, against NO_x and anthropogenic VOC emissions.

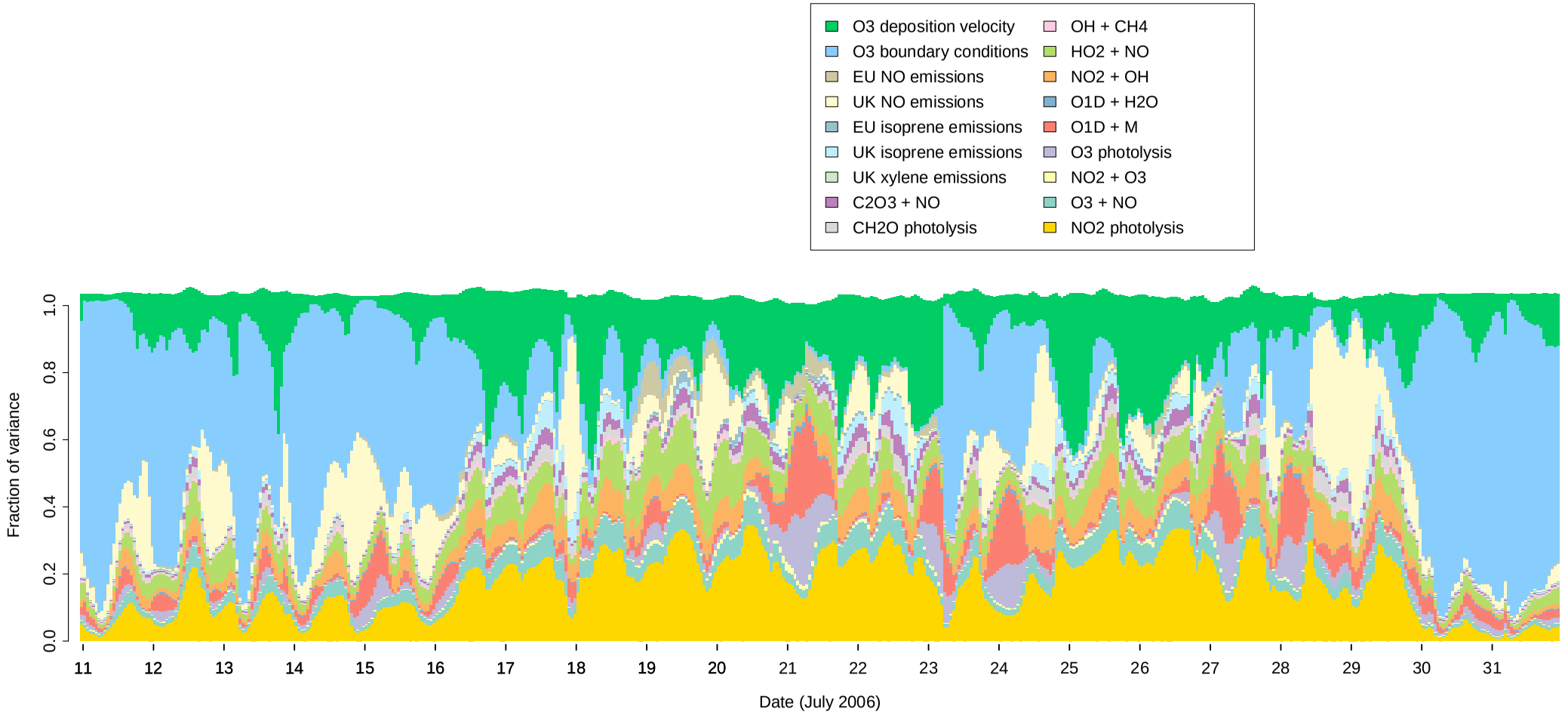


Figure A.7: Time series of Main effects on modelled ozone concentrations at Ladybower, 11th-31st July

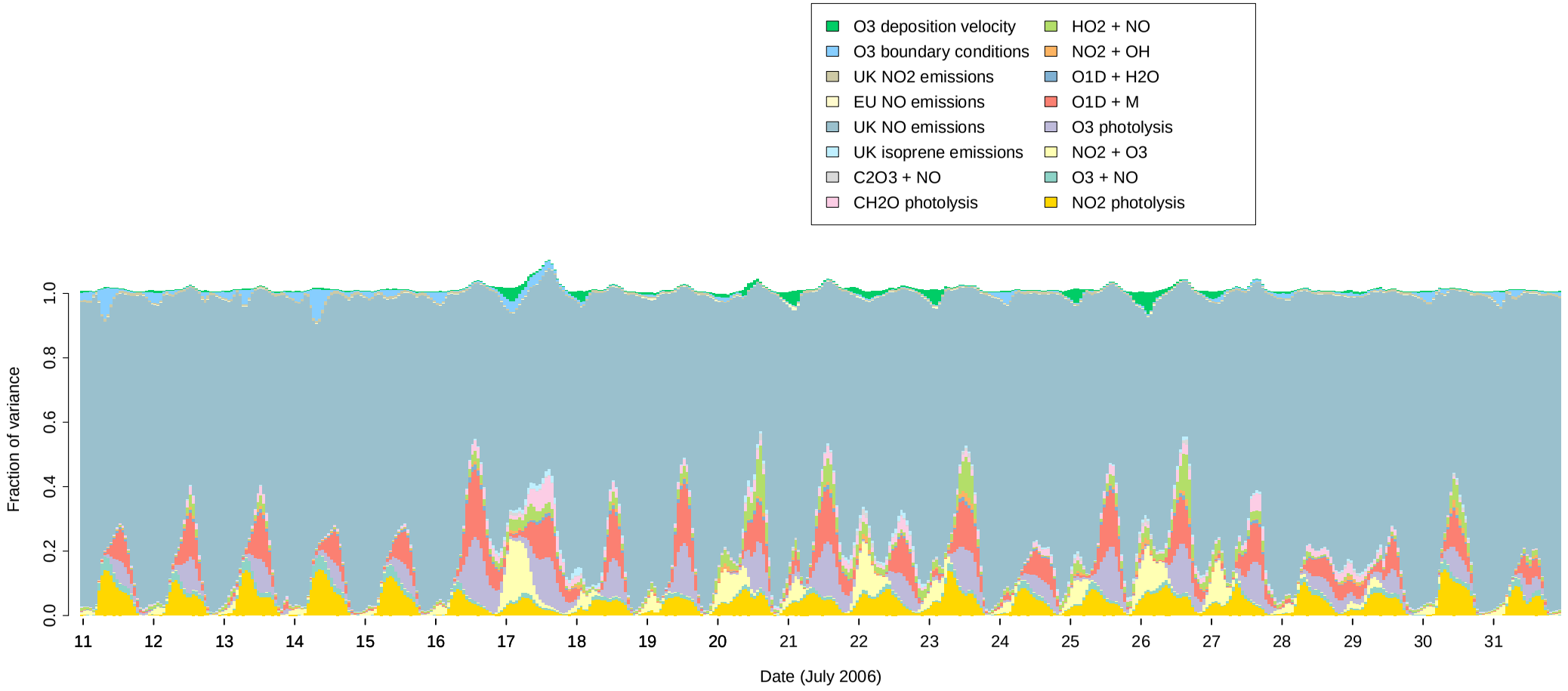


Figure A.8: Time series of Main effects on modelled NO_2 concentrations at Ladybower, 11th-31st July

A.3 Manchester Picadilly

The monitoring station is within a self contained air conditioned housing in the west-end of central Manchester in a pedestrianised zone approx 3 metres from (electric) tramline. The nearest major road, the A5103 Portland Street is approximately 200 metres from the station. The surrounding area is generally open with commercial property.

Environment Type: Urban Background

Altitude: 45 m

Latitude/Longitude: 53.481520, -2.237881

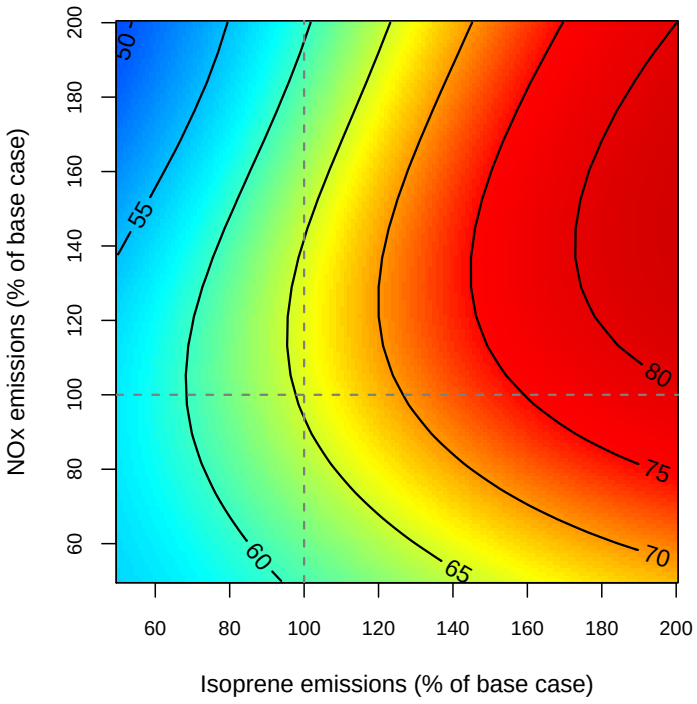


Figure A.9: Ozone concentrations at Manchester Picadilly, labelled in ppb on the contours, against NO_x and isoprene emissions.

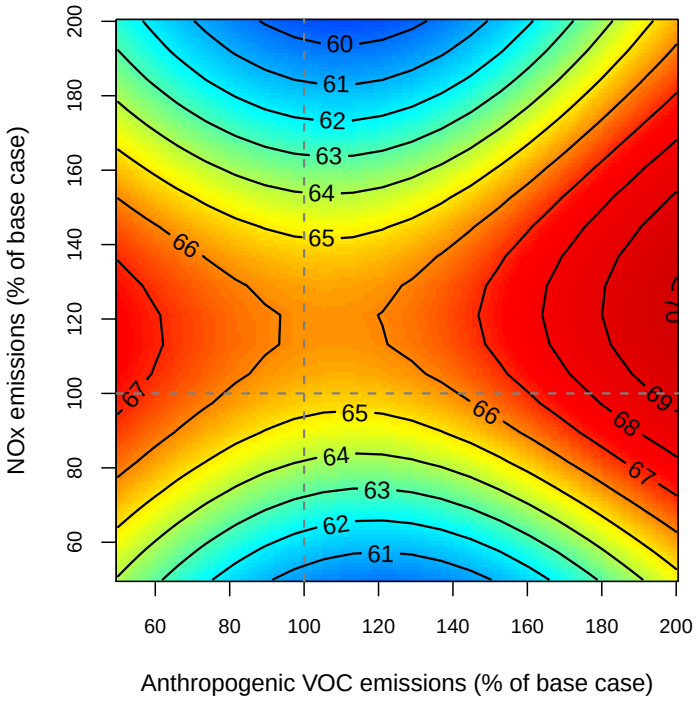


Figure A.10: Ozone concentrations at Manchester Picadilly, labelled in ppb on the contours, against NO_x and anthropogenic VOC emissions.

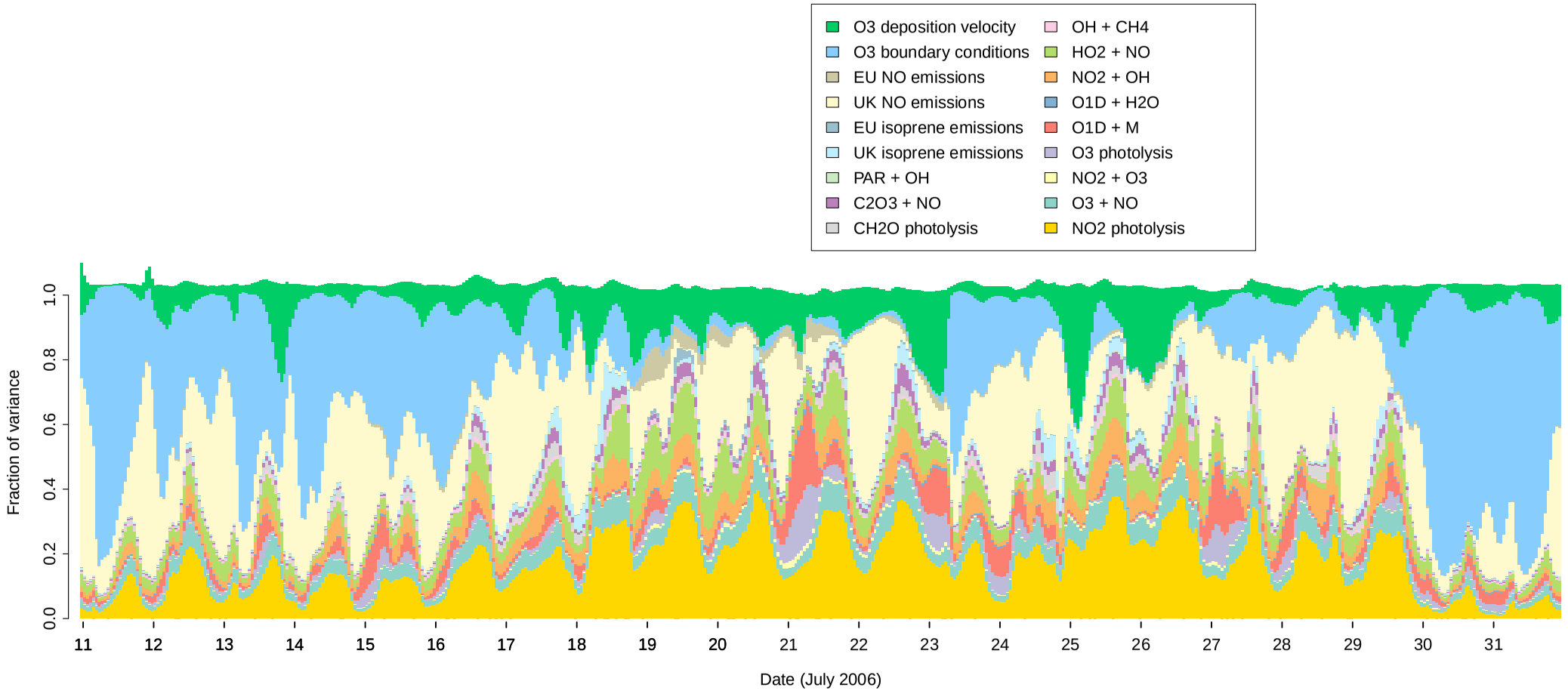


Figure A.11: Time series of Main effects on modelled ozone concentrations at Manchester Picadilly, 11th-31st July

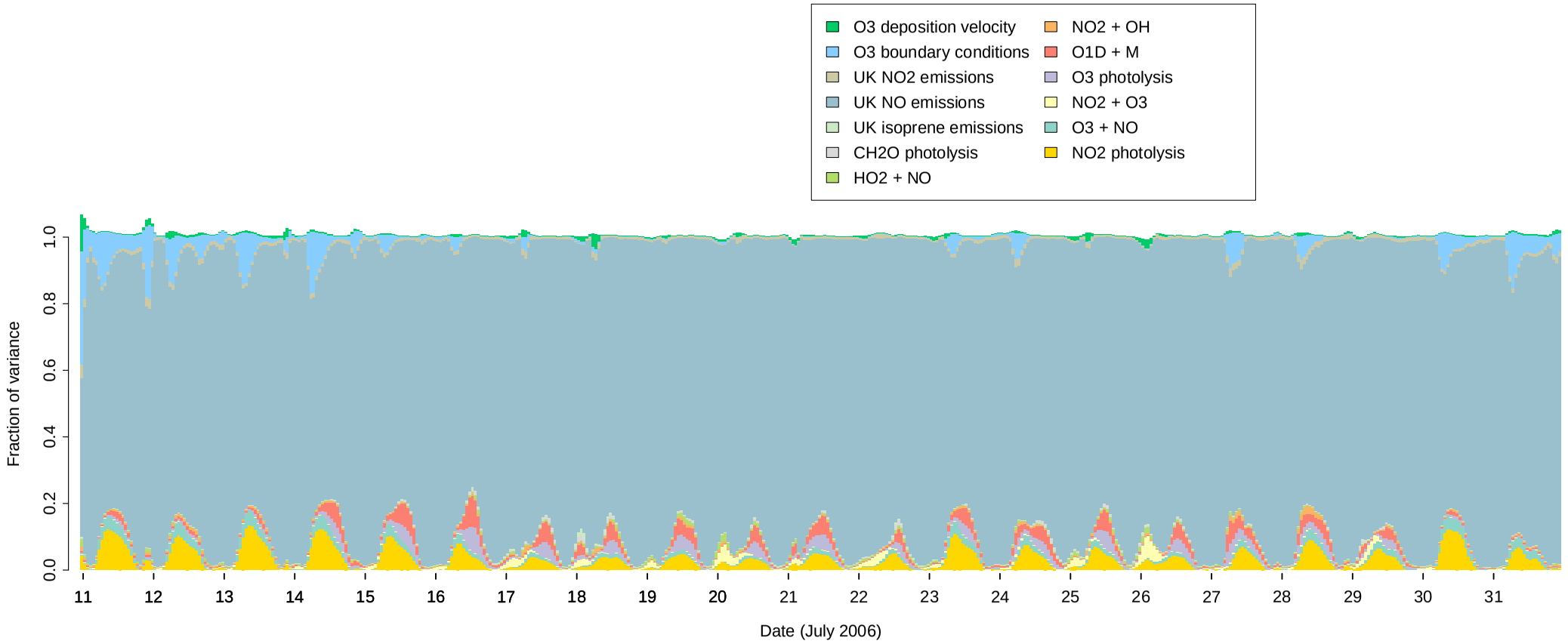


Figure A.12: Time series of Main effects on modelled NO_2 concentrations at Manchester Picadilly, 11th-31st July

A.4 Strath Vaich

The monitoring station is within a self-contained, air conditioned housing located on remote moorland approximately 500 metres from the nearest inhabited dwellings. The nearest road is approximately 150 metres from the site and used for access only. The surrounding area is open and remote. The site is at the SE corner of a large body of open water - Loch Vaich.

Environment Type: Rural Background

Altitude: 270 m

Latitude/Longitude: 57.734456, -4.776583

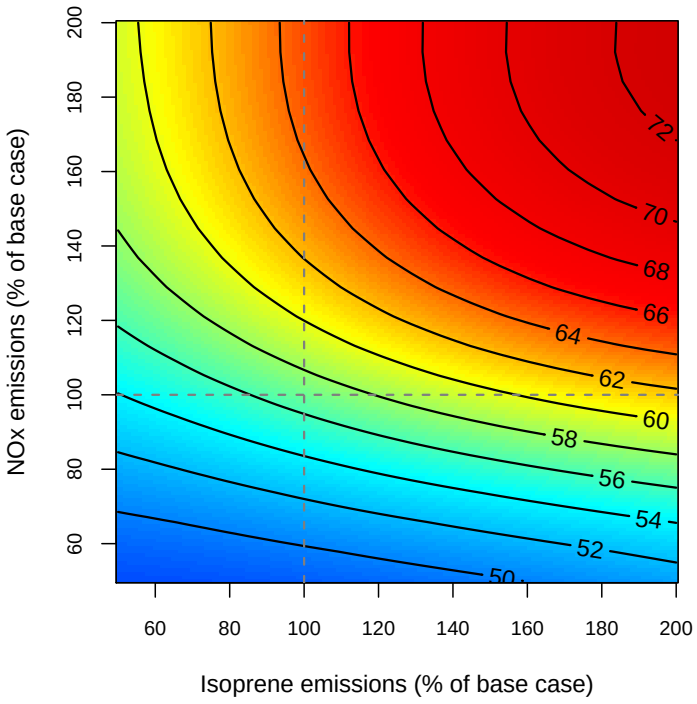


Figure A.13: Ozone concentrations at Strath Vaich, labelled in ppb on the contours, against NO_x and isoprene emissions.

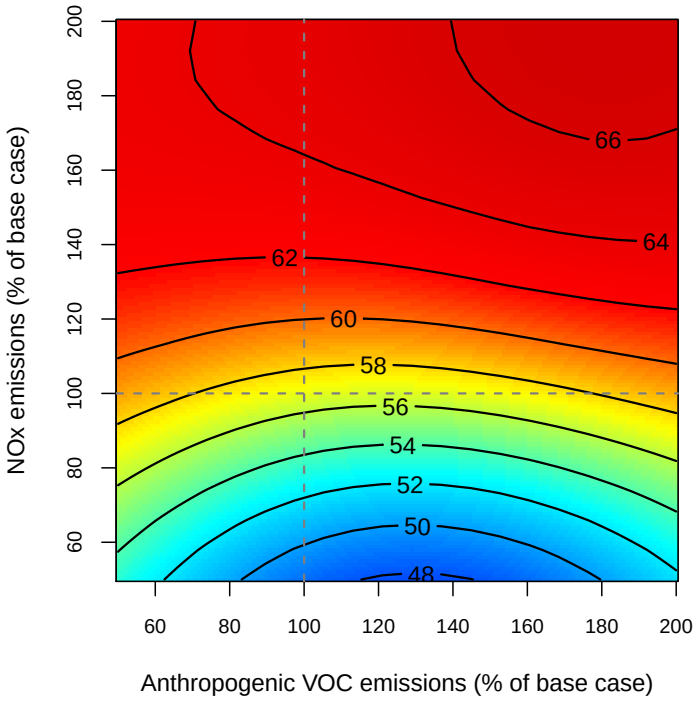


Figure A.14: Ozone concentrations at Strath Vaich, labelled in ppb on the contours, against NO_x and anthropogenic VOC emissions.

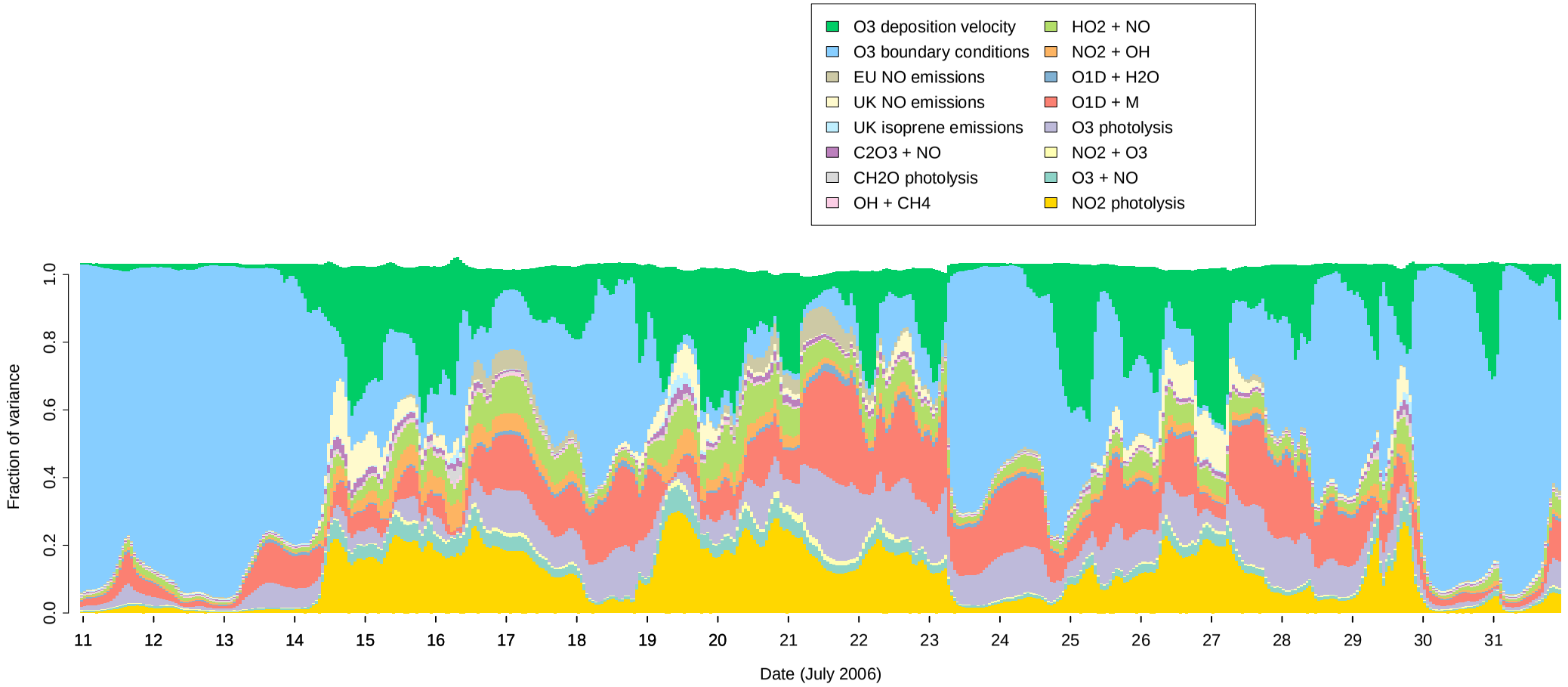


Figure A.15: Time series of Main effects on modelled ozone concentrations at Strath Vaich, 11th-31st July

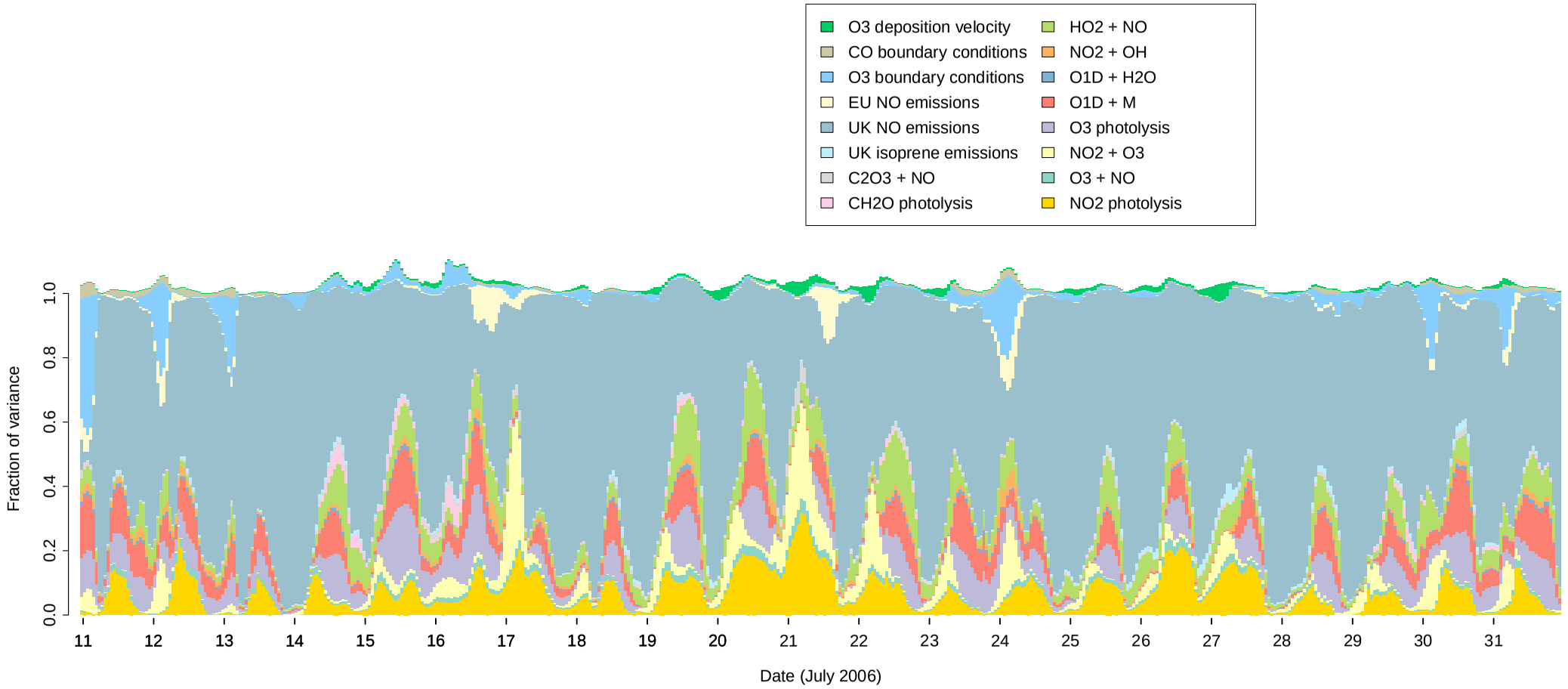


Figure A.16: Time series of Main effects on modelled NO_2 concentrations at Strath Vaich, 11th-31st July

A.5 Yarner Wood

The monitoring station is within a self-contained, air-conditioned housing located in a rural landscape. The nearest road is an access road approximately 60 metres distance. The surrounding area is partly wooded, partly open and comprises rural heath land.

Environment Type: Rural Background

Altitude: 119 m

Latitude/Longitude: 50.597600, -3.716510

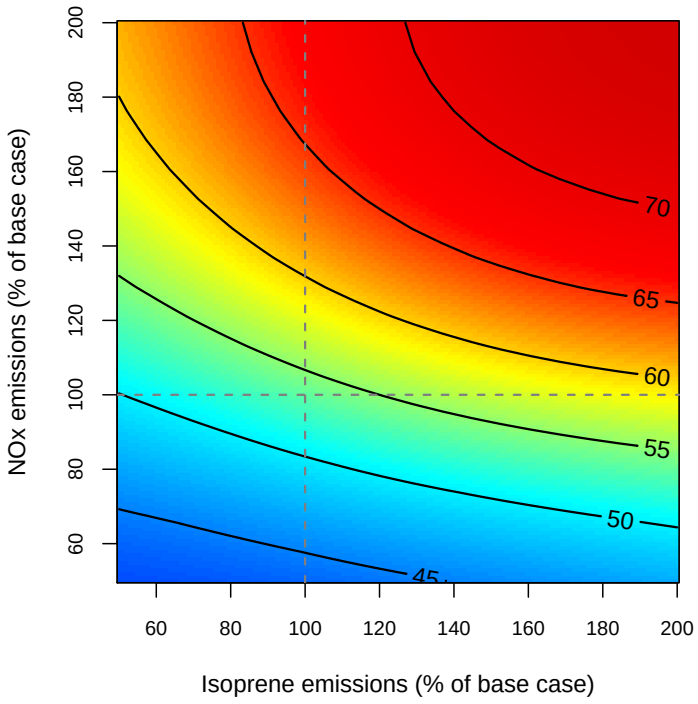


Figure A.17: Ozone concentrations at Yarner Wood, labelled in ppb on the contours, against NO_x and isoprene emissions.

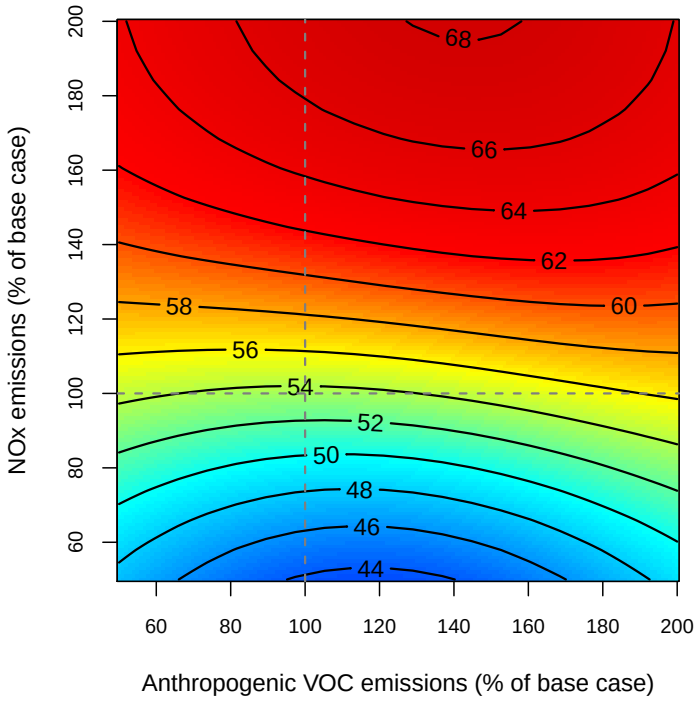


Figure A.18: Ozone concentrations at Yarner Wood, labelled in ppb on the contours, against NO_x and anthropogenic VOC emissions.

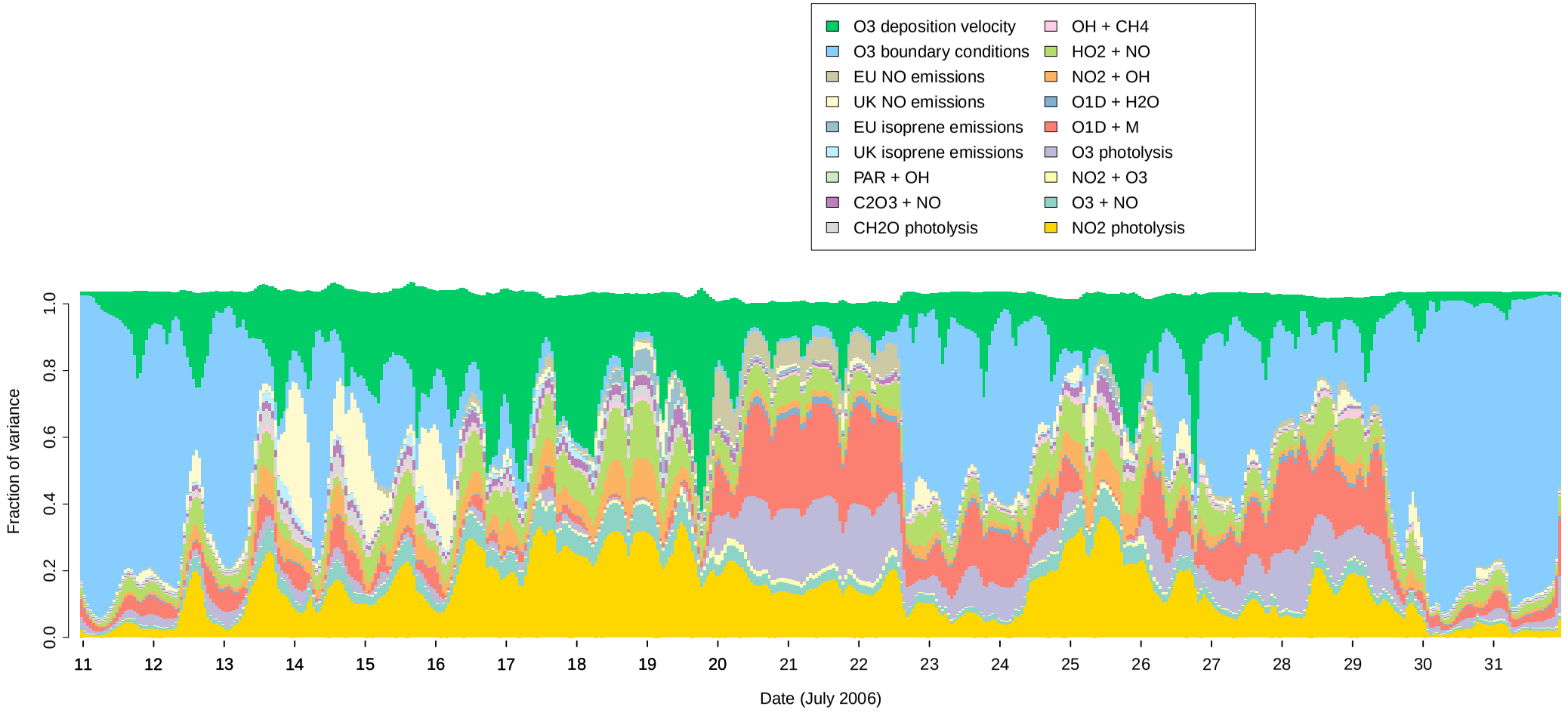


Figure A.19: Time series of Main effects on modelled ozone concentrations at Yarner Wood, 11th-31st July

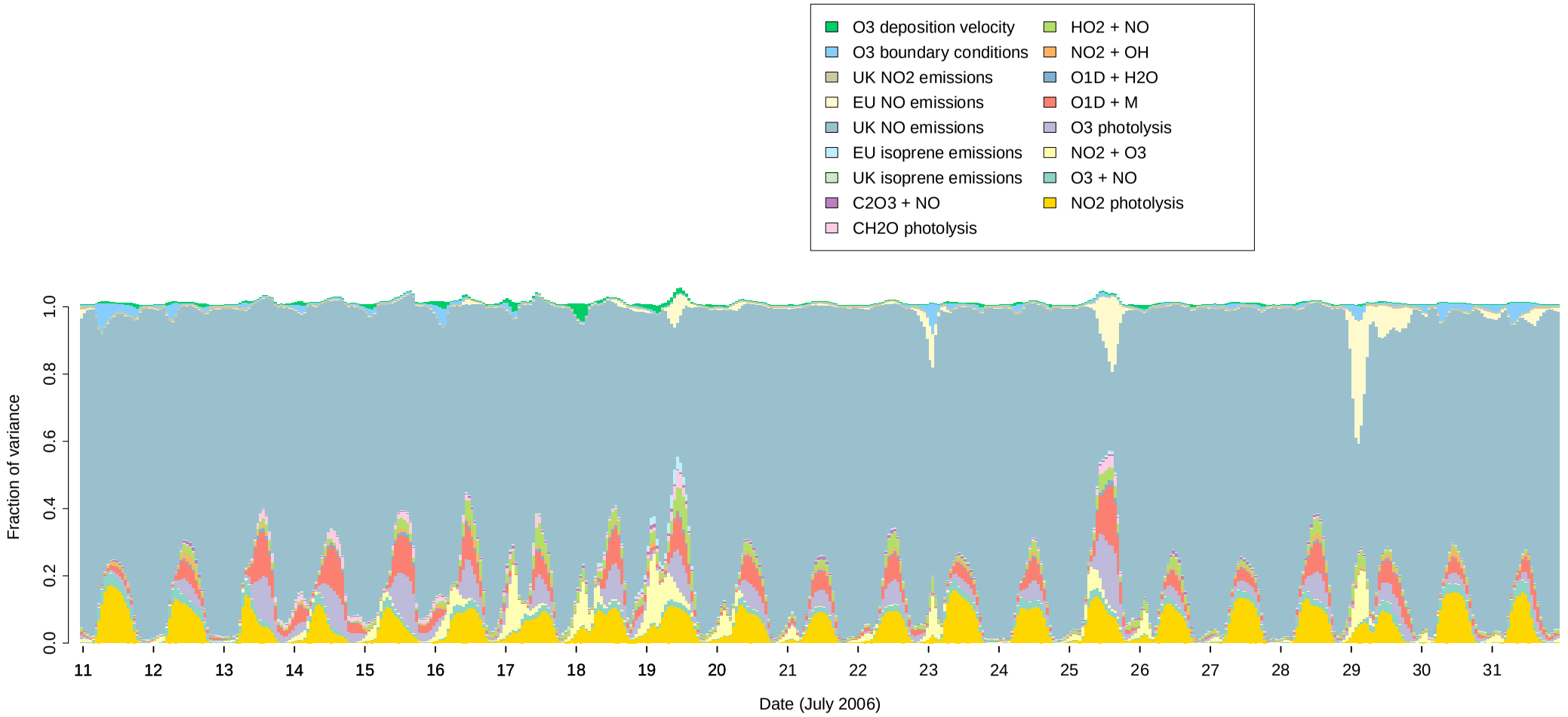


Figure A.20: Time series of Main effects on modelled NO_2 concentrations at Yarner Wood, 11th-31st July

Appendix B

Code Listings

The following listings represent part of the code written during the course of the PhD and are not intended to be a complete 'cut and paste' system, but must be used with CMAQ run scripts and source code. They are intended to assist the reader with some knowledge of R and Linux shell scripting to carry out similar analyses.

B.1 Automated system for CMAQ runs

B.1.1 Master run script

```
#! /bin/bash

read -a depV < depV.txt

declare -i n=0

while read line
do
RXlist[n]=$line
n=$((++n))
done < RXlist.txt

for ((j=0; j<=11; j++))
do

for ((i=1; i<=48; i++))
do

  k=$((i+j*48))
```

```
  l=$((k-1))

  cat RXDTstart.EXT RXdir/${RXlist[l]} RXDTend.EXT > /home/
     andrew/CMAQ_4.7.1_81/data${i}/cb05cl_ae5_aq/RXDT.EXT

  sed -e "26s/      1.0 /${depV[l]}/" /home/andrew/CMAQ_4.7.1_81
     /models/include/release/cb05cl_ae5_aq/GC_DEPV.EXT > /home/
     andrew/CMAQ_4.7.1_81/data${i}/cb05cl_ae5_aq/GC_DEPV.EXT

  cd cctm
  rm -rf BLD_e1a
  export M3DATA=/home/andrew/CMAQ_4.7.1_81/data${i}
  ./bldit.many.cctms >& build.log
  mv CCTM_e1a_Linux2_x86_64pg $M3DATA
  cd ..

  rm /home/andrew/CMAQ_4.7.1_81/data${i}/emis/2006/*
  cp /home/andrew/CMAQ_4.7.1_81/data/emis/2006/* /home/andrew/
     CMAQ_4.7.1_81/data${i}/emis/2006/
  /shared/R-3.0.2/bin/Rscript --vanilla changeDomainEmis.R i=$i
       k=$k



  rm /home/andrew/CMAQ_4.7.1_81/data${i}/raw/bcon/*
  /shared/R-3.0.2/bin/Rscript --vanilla readProfile.R i=$i k=$k

  cp -s /home/andrew/STORE/cgrids.2006201/data${k}/* /home/
     andrew/CMAQ_4.7.1_81/data${i}/cctm

done

ssh ctm00 "~/CMAQ_4.7.1_81/scripts/run.series1 > /dev/null 2>&1
    "

ssh ctm01 "~/CMAQ_4.7.1_81/scripts/run.series2 > /dev/null 2>&1
    "

ssh ctm02 "~/CMAQ_4.7.1_81/scripts/run.series3 > /dev/null 2>&1
    "
```

```
ssh ctm03 "~/CMAQ_4.7.1_81/scripts/run.series4 > /dev/null 2>&1
    "

for ((i=1; i<=48; i++))
do
  while [ ! -e /home/andrew/CMAQ_4.7.1_81/data${i}/cctm/
      CCTM_e1a_Linux2_x86_64pg.EU09.CGRID.2006212.ncf ]
  do
    sleep 5
  done
done

sleep 120

for ((i=1; i<=48; i++))
do

  k=$((i+j*48))
  mv /home/andrew/CMAQ_4.7.1_81/data${i}/cctm/
      CCTM_e1a_Linux2_x86_64pg.EU??.ACONC* /home/andrew/DATA/
      data${k}

  for name in /home/andrew/CMAQ_4.7.1_81/data${i}/cctm/
      CCTM_e1a_Linux2_x86_64pg.EU09.DRYDEP*
  do
    ncks -O -a -v TFLAG,NO2,NO,O3 $name $name
  done

  mv /home/andrew/CMAQ_4.7.1_81/data${i}/cctm/
      CCTM_e1a_Linux2_x86_64pg.EU09.DRYDEP* /home/andrew/DATA/
      data${k}
  mv /home/andrew/CMAQ_4.7.1_81/data${i}/cctm/
      CCTM_e1a_Linux2_x86_64pg.EU??.CGRID.2006130.ncf /home/
      andrew/DATA/data${k}

  rm /home/andrew/CMAQ_4.7.1_81/data${i}/cctm/*
  rm /home/andrew/CMAQ_4.7.1_81/data${i}/icon/*
  rm /home/andrew/CMAQ_4.7.1_81/data${i}/bcon/*
  rm /home/andrew/CMAQ_4.7.1_81/data${i}/raw/bcon/*
  rm /home/andrew/CMAQ_4.7.1_81/data${i}/cb05cl_ae5_aq/RXDT.EXT
```

```
  rm /home/andrew/CMAQ_4.7.1_81/data${i}/cb05cl_ae5_aq/GC_DEPV.
     EXT
  rm /home/andrew/CMAQ_4.7.1_81/data${i}/
     CCTM_e1a_Linux2_x86_64pg

done

done
```

B.1.2 R code to modify emissions files

```
rm(list=ls())

library(ncdf)

args<-commandArgs(TRUE)
eval(parse(text=args))

hyper<-read.table('hyper.txt')

setwd(paste('~/CMAQ_4.7.1_81/data',i,'/emis/2006',sep=''))

files<-list.files()
eu9files<-files[grep('eu9',files)]
eu27files<-files[grep('eu27',files)]
eu81files<-files[grep('eu81',files)]

vars<-c('ALD2','ALDX','BENZENE','CH4','CO','ETH','ETHA','ETOH',
    'FORM','HCL','HONO','IOLE',
        'ISOP','MEOH','NASN','NH3','NO','NO2','NR','OLE','PAR','
            SESQ','SO2','SULF','TERP',
        'TOL','XYL')

n<-12

for (var in vars)
   {
   n<-n+1
   for (file in eu9files)
      {
      nc<-open.ncdf(file,write=T)
```

```
        store<-get.var.ncdf(nc,var,c(1,1,1,1),c(-1,-1,-1,-1))
        store<-store*hyper[k,n]
        put.var.ncdf(nc,var,store,c(1,1,1,1),c(-1,-1,-1,-1))
        close.ncdf(nc)
        }
    n<-n+1
    for (file in eu27files)
        {
        nc<-open.ncdf(file,write=T)
        store<-get.var.ncdf(nc,var,c(1,1,1,1),c(-1,-1,-1,-1))
        store<-store*hyper[k,n]
        put.var.ncdf(nc,var,store,c(1,1,1,1),c(-1,-1,-1,-1))
        close.ncdf(nc)
        }
    for (file in eu81files)
        {
        nc<-open.ncdf(file,write=T)
        store<-get.var.ncdf(nc,var,c(1,1,1,1),c(-1,-1,-1,-1))
        store<-store*hyper[k,n]
        put.var.ncdf(nc,var,store,c(1,1,1,1),c(-1,-1,-1,-1))
        close.ncdf(nc)
        }
    }
```

B.1.3 R code to modify boundary conditions files

```
rm(list=ls())

library(ncdf)
library(MASS)

args=(commandArgs(TRUE))
eval(parse(text=args))

infile<-'Combined_GEMS_20060628_n_bc_Profile_V9.txt'

north<-read.table(infile,skip=6,nrows=105)
east<-read.table(infile,skip=113,nrows=105)
south<-read.table(infile,skip=220,nrows=105)
west<-read.table(infile,skip=327,nrows=105)
```

```
hyper<-read.table('hyper.txt')

setwd(paste('~/CMAQ_4.7.1_81/data',i,'/raw',sep=''))

#ic<-(north[,-1]+east[,-1]+south[,-1]+west[,-1])/4
#ic<-cbind(north[,1],ic)

var<-c(1:17,24:30,33:45,48:50)

for (n in var)
    {
    m<-var[n]
    north[m,-1]<-north[m,-1]*hyper[k,28+n]
    east[m,-1]<-east[m,-1]*hyper[k,28+n]
    south[m,-1]<-south[m,-1]*hyper[k,28+n]
    west[m,-1]<-west[m,-1]*hyper[k,28+n]
    }

north[,-1]<-signif(north[,-1],digits=4)
east[,-1]<-signif(east[,-1],digits=4)
south[,-1]<-signif(south[,-1],digits=4)
west[,-1]<-signif(west[,-1],digits=4)

head<-data.frame(initial=c(0,0,8,2006121),conditions=c
    (0,0,105,00),and=c(0,0,
                       1.00,0),boundary=c(0,0,0.99,0),conds=c
                          (0,0,0.96,0),
                         file=c(0,0,0.85,0),header=c(0,0,0.67,0),
                            For=c(0,0,0.46,0),
                        CMAQ=c(0,0,0.26,0),HA=c(0,0,0.10,0),HAA=c
                            (0,0,0.00,0))

#ic_file<-'icon/ic_profile_v9.dat'
bc_file<-'bcon/bc_profile_v9.dat'

#write.table(head,ic_file,row.names=F,quote=F)

#write.table(prettyNum(ic),ic_file,row.names=F,col.names=F,
#                       quote=1,sep='    ',append=T)

n<-' North'
```

```
e<-' East'
s<-' South'
w<-' West'

write.table(head,bc_file,row.names=F,quote=F)

write.table(n,bc_file,row.names=F,col.names=F,quote=F,append=T)

write.table(prettyNum(north),bc_file,row.names=F,col.names=F,
                    quote=1,sep='   ',append=T)

write.table(e,bc_file,row.names=F,col.names=F,quote=F,append=T)

write.table(prettyNum(east),bc_file,row.names=F,col.names=F,
                    quote=1,sep='   ',append=T)

write.table(s,bc_file,row.names=F,col.names=F,quote=F,append=T)

write.table(prettyNum(south),bc_file,row.names=F,col.names=F,
                    quote=1,sep='   ',append=T)

write.table(w,bc_file,row.names=F,col.names=F,quote=F,append=T)

write.table(prettyNum(west),bc_file,row.names=F,col.names=F,
                    quote=1,sep='   ',append=T)
```

B.1.4 FORTRAN code to modify RXDT.EXT source code files

```
program react3

implicit none

double precision :: A(156),B(156),C(156),H(12)
integer :: RX(12)=(/1,3,7,9,10,11,28,30,66,74,87,112/)
integer :: m(5)=(/1,2,3,4,5/)
integer :: n(5),i,j,k
character*31 :: str
character*3 :: num
str="O+1+2+3+4+5+6+7+8+9+O+1+2+3+4+5"
B(1:156)=1.0
```

```
open(1,file="RX1")
read(1,*)A
close(1)

do i=1,576

   open(2,file="hyper.txt")
   read(2,*)H

   do k=1,12

   B(RX(k))=H(k)

   end do

   write(num,300) i
   300 format(I3.3)
   open(3,file='RX.'//num//'.out')

   C=A*B

   do j=1,31

      n=m+((j-1)*5)

      write (3,100), str(j:j),C(n(1)), C(n(2)), C(n(3)), C(n(4)),
          C(n(5))
      100 format(5x,A1,5x,D10.4,',  ',D10.4,',  ',D10.4',  ',D10.4',
          ',D10.4',')

   end do

   write (3,200), C(156)
   200 format(5x,'+',5x,D10.4,', 1.0000D+00, 1.0000D+00, 2.3000D
      -11, 1.6300D-14,')

   close(3)

end do
```

```
close(2)

end program react3
```

B.1.5 run.series1 script

```
#! /bin/bash

export M3HOME=/home/andrew/CMAQ_4.7.1_81
export M3LIB=$M3HOME/lib
cd $M3HOME/scripts

for ((i=1; i<=12; i++))
do

   export M3DATA=/home/andrew/CMAQ_4.7.1_81/data${i}

   ./run.master10 &

done
```

B.1.6 run.master10 script

```
#! /bin/bash

cd $M3HOME/scripts

cd icon
./run.iconDAY1outer

cd ../bcon
./run.bcon_profile

cd ../cctm
./run.cctmOuterFirst
./run.cctmOuter

cd ../bcon2
./run.bconMiddle
```

```
cd ../icon2
./run.iconMiddle

rm $M3DATA/cctm/*AEROVIS*
rm $M3DATA/cctm/*.CONC*
rm $M3DATA/cctm/*WETDEP*

cd ../cctm
./run.cctmMiddle

cd ../bcon2
./run.bconInner

cd ../icon2
./run.iconInner

rm $M3DATA/cctm/*AEROVIS*
rm $M3DATA/cctm/*.CONC*
rm $M3DATA/cctm/*WETDEP*

cd ../cctm
./run.cctmInner
```

B.2 R script to apply Morris' method

The file `morris.txt` must be produced beforehand with the function `morris` from the *sensitivity* package and used with the scripts in section B.1 to perform the CMAQ training runs.

```
rm(list=ls())

library(ncdf)
library(sensitivity)

morrisDesign<-read.table('morris.txt',header=F)
colnames(morrisDesign)<-c('ALD2','ALDX','BENZENE','CH4','CO',
            'ETH','ETHA','ETOH','FORM','HCL','HONO','IOLE',
            'ISOP','MEOH','NASN','NH3','NO','NO2','NR','OLE',
            'PAR','SESQ','SO2','SULF','TERP','TOL','XYL',
            'NO2.b','NO.b','O3.b','HNO3.b','H2O2.b',HCHO.b',
            'PAN.b','HC5.b','ETH.b','CO.b','ALD.b','ISO.b',
```

```
            'OLT.b', 'OLI.b', 'TOL.b', 'XYL.b', 'SO2.b', 'OP1.b',
            'OP2.b', 'PAA.b', 'ORA1.b', 'ORA2.b', 'NH3.b', 'HC3.b',
            'HC8.b', 'OL2.b', 'ACO3.b', 'TPAN.b', 'HONO.b', 'HNO4.b',
            'KET.b', 'GLY.b', 'MGLY.b', 'DCB.b', 'ONIT.b', 'CSL.b',
            'MACR.b', 'MVK.b', 'BENZENE.b')

morrisDesign<-morrisDesign[1:670,]

morrisScreen<-morris(model=NULL,factors=66,r=10,design=list(
    type='oat',
                         levels=7,grid.jump=3),binf=0.5,bsup=2)

morrisScreen$X<-as.matrix(morrisDesign)
sites<-read.csv('MIE_grid_point.csv')

muGt2ppb<-NULL
muGt1.5ppb<-NULL
muGt1ppb<-NULL
mu.5to1ppb<-NULL

results<-data.frame(sites$Site_ID,morrisDesign[1:22,])

for (j in 1:22)
    {
    col<-sites$COL[j]
    row<-sites$ROW[j]
    output<-0.0

    for (i in 1:670)
        {
        setwd(paste('./data',i,sep=''))
        nc<-open.ncdf('CCTM_e1a_Linux2_x86_64pg.EU09.ACONC
            .2006200.ncf')
        output[i]<-get.var.ncdf(nc,'O3',c(col,row,1,17),c
            (1,1,1,1))
        close.ncdf(nc)
        setwd('..')
        }

    tell(morrisScreen,output)
    plot(morrisScreen,identify=T,cex.lab=1.5,cex.axis=1.2,col='
```

```
    red')

  mu<-apply(morrisScreen$ee, 2, mean)
  mu.star<-apply(morrisScreen$ee,2,
                 function(morrisScreen) mean(abs(morrisScreen))
                   )
  sigma<-apply(morrisScreen$ee, 2, sd)

  results[j,2:67]<-mu.star

  results<-results[order(results$mu.star,decreasing=T),]

  index<-mu.star>=.002
  muGt2ppb<-append(muGt2ppb,names(mu.star[index]))
  index<-mu.star>=.0015
  muGt1.5ppb<-append(muGt1.5ppb,names(mu.star[index]))
  index<-mu.star>=.001
  muGt1ppb<-append(muGt1ppb,names(mu.star[index]))
  index2<-mu.star>=.0005 & mu.star<.001
  mu.5to1ppb<-append(mu.5to1ppb,names(mu.star[index2]))
  muGt1ppb<-unique(muGt1ppb)
  muGt1.5ppb<-unique(muGt1.5ppb)
  muGt2ppb<-unique(muGt2ppb)
  mu.5to1ppb<-unique(mu.5to1ppb)
  }
```

B.3 R code to test the BACCO and DiceKriging packages

The file `hyper.txt` must be produced beforehand with the *BACCO* package or another package capable of producing LHS and used with the scripts in section B.1 to perform the CMAQ training runs.

```
rm(list=ls())

library(ncdf)
library(DiceKriging)
library(BACCO)

hyper<-read.table('hyper.txt',header=F)
```

```
colnames(hyper)<-c('ETH.i','ETH.o','ISOP.i','ISOP.o','NO.i','NO
    .o','NO2.i','NO2.o','PAR.i','PAR.o','XYL.i','XYL.o','CO.i','
    CO.o','bO3','bCO')

output<-0.0

for (i in 1:336)
    {
    setwd(paste('./data',i,sep=''))
    nc<-open.ncdf('CCTM_e1a_Linux2_x86_64pg.EU09.ACONC.2006200.
        ncf')
    output[i]<-get.var.ncdf(nc,'O3',c(41,17,1,17),c(1,1,1,1))
    close.ncdf(nc)
    setwd('..')
    }

sample<-sample(1:336,36)
trainHyper<-as.matrix(hyper[-sample,])
trainOut<-output[-sample]
testHyper<-as.matrix(hyper[sample,])
testOut<-output[sample]

model<-km(design=trainHyper,response=trainOut)
Dice<-predict.km(model,testHyper,'UK')$mean

scales.optim<-optimal.scales(trainHyper,scales.start=rep(1,ncol
    (trainHyper)),trainOut)

BACC<-NULL

for (i in 1:36)
  {
  BACC[i]<-interpolant(testHyper[i,],trainOut,trainHyper,give.
      full.list=T,scales=scales.optim)$mstar.star
  }

png(file='diceVSbacco.png',width=500,height=500)

par(mar=c(5.1,4.1,1.1,2.1))

plot(testOut,Dice,xlab='CMAQ O3 (ppm)',ylab='Dice / BACCO O3 (
```

```
    ppm)',col='red',cex.lab=1.3,cex.axis=1.2,pch=16)
points(testOut,BACC, pch=16,col='blue')
abline(0,1,col='black',lwd=1.5,lty=1)

legend(x=.075,y=.105,legend=c('Dice','BACCO'),pch=c(16,16),col=
    c('red','blue'),cex=1.3)

dev.off()
```

B.4 R code to emulate CMAQ and produce FAST time series

```
rm(list=ls())

library(ncdf)
library(DiceKriging)
library(sensitivity)
library(RColorBrewer)
library(truncnorm)

hyper<-read.table('hyper.txt',header=F)

colnames(hyper)<-c('NO2 phot','O3+NO','NO2+O3','O3 phot',
                   'O1D+M','O1D+H2O','NO2+OH','HO2+NO',
                   'OH+CH4','CH2O phot','C2O3+NO','PAR+OH',
                   'UK ISOP','EU ISOP','UK NO','EU NO',
                   'UK NO2','EU NO2','UK PAR','EU PAR',
                   'UK XYL','EU XYL','UK ETH','EU ETH',
                   'UK CO','EU CO','UK OLE','EU OLE',
                   'O3 BC','CO BC','O3 dep V')

days<-c(192:193)
ndays<-length(days)
nhours<-ndays*24
TS<-matrix(0.0,ncol(hyper),nhours)
rownames(TS)<-colnames(hyper)

for (k in 1:length(days))
  {
```

```
for (j in 1:24)
    {
    output<-0.0

    for (i in 1:576)
        {
        setwd(paste('./data',i,sep=''))
        nc<-open.ncdf(paste('CCTM_e1a_Linux2_x86_64pg.EU09.ACONC
            .2006',
                                days[k],'.ncf',sep=''))
        output[i]<-get.var.ncdf(nc,'O3',c(41,17,1,j),c(1,1,1,1))
        close.ncdf(nc)
        setwd('..')
        }

    model<-km(design=hyper,response=output)

    kriging.mean<-function(Xnew,m)
        {
         predict.km(m,Xnew,'UK',se.compute=F,checkNames=F)$mean
        }

fastSA<-fast99(model=kriging.mean,factors=colnames(hyper),n
    =5000,
                q=c('qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qtruncnorm',
                    'qunif',
                    'qunif',
                    'qunif',
                    'qunif',
                    'qunif',
```

```
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif',
        'qunif'),
    q.arg=list(list(a=0.5,b=2.0,mean=1.0,sd=0.2),   #
        NO2 phot
               list(a=0.5,b=2.0,mean=1.0,sd=0.1),   #
                   O3+NO
               list(a=0.5,b=2.0,mean=1.0,sd=0.15),  #
                   NO2+O3
               list(a=0.5,b=2.0,mean=1.0,sd=0.3),   #
                   O3 phot
               list(a=0.5,b=2.0,mean=1.0,sd=0.5),   #
                   O1D+M
               list(a=0.5,b=2.0,mean=1.0,sd=0.08),  #
                   O1D+H2O
               list(a=0.5,b=2.0,mean=1.0,sd=0.4),   #
                   NO2+OH
               list(a=0.5,b=2.0,mean=1.0,sd=0.3),   #
                   HO2+NO
               list(a=0.5,b=2.0,mean=1.0,sd=0.1),   #
                   OH+CH4
               list(a=0.5,b=2.0,mean=1.0,sd=0.4),   #
                   CH2O phot
               list(a=0.5,b=2.0,mean=1.0,sd=0.5),   #
                   C2O3+NO
               list(a=0.5,b=2.0,mean=1.0,sd=0.3),   #
                   PAR+OH
               list(min=0.7,max=1.3),
               list(min=0.7,max=1.3),
               list(min=0.7,max=1.3),
```

```
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.7,max=1.3),
                                    list(min=0.5,max=1.5),
                                    list(min=0.5,max=1.5),
                                    list(min=0.7,max=1.3)),m=model)

    TS[,(k-1)*24+j]<-1-(fastSA$Dt/fastSA$V)
    }
}


index<-NULL

for (k in 1:nrow(TS))
    {
    if (max(TS[k,])<0.01) index[k]<-F else index[k]<-T
    }

TS<-TS[index,]

for (l in 1:48)
    {
    hourSum<-sum(TS[,l])
    TS[,l]<-TS[,l]/hourSum
    }

colnames(TS)<-1:48

barplot(TS,space=0,legend.text=T,col=brewer.pal(nrow(TS),'
    Pastel1'),border=NA,
```

```
        args.legend=c(x=8,y=0.8),xlab='Hour',ylab='Fraction of
            Variance',
        cex.axis=1.2,cex.lab=1.5)
```

B.5 R code to produce the plots in section 4.3

```
rm(list=ls())

library(ncdf)
library(DiceKriging)
library(truncnorm)
library(reshape)
library(scales)

n<-10000
factor<-c(16,18)

hyper<-read.table('hyper.txt',header=F)

colnames(hyper)<-c('R1','R3','R7','R9','R10','R11','R28','R30',
                   'R66','R74','R87','R112','ISOP.i','ISOP.o',
                   'NO.i','NO.o','NO2.i','NO2.o','PAR.i',
                   'PAR.o','XYL.i','XYL.o','ETH.i','ETH.o',
                   'CO.i','CO.o','OLE.i','OLE.o','O3.b','CO.b',
                   'depV')

output<-0.0

for (i in 1:576)
   {
   setwd(paste('./data',i,sep=''))
   nc<-open.ncdf('CCTM_e1a_Linux2_x86_64pg.EU09.ACONC.2006200.
       ncf')
   output[i]<-get.var.ncdf(nc,'O3',c(50,14,1,17),c(1,1,1,1))
   close.ncdf(nc)
   setwd('..')
   }

output<-output*1000
```

```
model<-km(design=hyper,response=output)

plot(model)

kriging.mean<-function(Xnew,m)
    {
     predict.km(m,Xnew,'UK',se.compute=F,checkNames=F)$mean
    }

means<-rep(1,31)
SDs<-c(.2,.1,.15,.3,.5,.08,.4,.3,.1,.4,.5,.3)
length(SDs)<-31
mins<-c(rep(.5,12),rep(.7,19))
maxes<-c(rep(2.0,12),rep(1.3,19))
dist<-data.frame(means,SDs,mins,maxes)
rownames(dist)<-colnames(hyper)
MCmatrix<-matrix(1,n,31)
colnames(MCmatrix)<-colnames(hyper)

for (i in 1:12)
  {
  MCmatrix[,i]<-rep(rtruncnorm(n/20,a=dist$mins[i],b=dist$maxes
      [i],
  mean=dist$means[i],sd=dist$SDs[i]),each=20)
  }

for (i in 13:31)
  {
  MCmatrix[,i]<-rep(runif(n/20,min=dist$mins[i],max=dist$maxes[
      i]),each=20)
  }

s<-rep(seq(0.5,2,length=20),(n/20))
MCmatrix[,factor]<-s

Y<-kriging.mean(MCmatrix,m=model)

pdf('KC1o3_noxemisEU_1600_19th.pdf',width=5,height=5)
par(mar=c(4.5,4.5,1,1))
```

```
plot(s[1:20],Y[1:20],type='l',ylim=c(min(Y),max(Y)),col='
    darkseagreen',
      ylab='Ozone (ppb)',xlab='NOx emissions (% of base case)',
         xaxt='n')

axis(side=1,at=c(.5,1,1.5,2),labels=c(50,100,150,200))

for(i in 1:500)
  {
  j<-(1:20)+(i*20)
  lines(s[1:20],Y[j],col='darkseagreen')
  }

out<-data.frame(s,Y)
pc5<-0.0
pc95<-0.0
pc25<-0.0
pc75<-0.0
m<-0.0

for(k in seq(0.5,2,length=20))
  {
  index<-out$s==k
  dist<-out$Y[index]
  pc5<-append(pc5,quantile(dist,.05))
  pc95<-append(pc95,quantile(dist,.95))
  pc25<-append(pc25,quantile(dist,.25))
  pc75<-append(pc75,quantile(dist,.75))
  m<-append(m,mean(dist))
  }

pc5<-pc5[-1]
pc95<-pc95[-1]
pc25<-pc25[-1]
pc75<-pc75[-1]
m<-m[-1]



x<-c(s[1:20],rev(s[1:20]))
y<-c(pc5,rev(pc95))
polygon(x,y,col=alpha('turquoise1',0.4),border=NA)
```

```
y<-c(pc25,rev(pc75))
polygon(x,y,col=alpha('darkslategray4',0.4),border=NA)

a<-matrix(1,20,length(means))
a[,factor]<-seq(0.5,2,length=20)
y<-kriging.mean(a,m=model)
lines(s[1:20],y,col='black',lwd=2)
lines(s[1:20],m,col='olivedrab2',lwd=2)

dev.off()

######## separate legend ##################

pdf('meLegend.pdf',width=5,height=5)
par(mar=c(4.5,4.5,1,1))

plot(s[1:20],Y[1:20],type='n',ylim=c(min(Y),max(Y)),
     col='darkseagreen',ylab='',xlab='',xaxt='n',
     yaxt='n')


legend(x=.62,y=117,
       legend=c('Emulator runs','Mean of emulator runs',
                '5th - 95th percentile of\n emulator runs',
                '25th - 75th percentile of\n emulator runs',
                'Emulator run with all other\n inputs at 100%')
                  ,
       col=c('darkseagreen','olivedrab2',alpha('turquoise1'
           ,0.4),
              alpha('darkslategray4',0.4),'black'),
       lty=1,lwd=c(2,3,10,10,3),bty='n',y.intersp=2,cex=1.2)

dev.off()
```

B.6 R code to produce the plots in section 4.4

```
rm(list=ls())

library(ncdf)
```

```
library(DiceKriging)
library(sensitivity)
library(plot3D)

devAskNewPage(ask=T)

hyper<-read.table('hyper.txt',header=F)

colnames(hyper)<-c('R1','R3','R7','R9','R10','R11','R28','R30',
                   'R66','R74','R87','R112','ISOP.i','ISOP.o',
                   'NO.i','NO.o','NO2.i','NO2.o','PAR.i',
                   'PAR.o','XYL.i','XYL.o','ETH.i','ETH.o',
                   'CO.i','CO.o','OLE.i','OLE.o','O3.b','CO.b',
                   'depV')

output<-0.0

for (i in 1:576)
    {
    setwd(paste('./data',i,sep=''))
    nc<-open.ncdf('CCTM_e1a_Linux2_x86_64pg.EU09.ACONC.2006200.
        ncf')
    output[i]<-get.var.ncdf(nc,'O3',c(40,41,1,17),c(1,1,1,1))
    close.ncdf(nc)
    setwd('..')
    }

model<-km(design=hyper,response=output)

kriging.mean<-function(Xnew,m)
    {
     predict.km(m,Xnew,'UK',se.compute=F,checkNames=F)$mean
    }

a<-matrix(data=rep(1,12400),nrow=400,ncol=31)
colnames(a)<-colnames(hyper)
s<-expand.grid(seq(0.5,2,length=20),seq(0.5,2,length=20))

#a[,13]<-s[,1] #isoprene
#a[,14]<-s[,1]
```

```
a[,15]<-s[,2] #NOx
a[,16]<-s[,2]
a[,17]<-s[,2]
a[,18]<-s[,2]

a[,19]<-s[,1] #anthrop. VOCs
a[,20]<-s[,1]
a[,21]<-s[,1]
a[,22]<-s[,1]
a[,23]<-s[,1]
a[,24]<-s[,1]
a[,27]<-s[,1]
a[,28]<-s[,1]

x=seq(.5,2,length=20)*100
y=seq(.5,2,length=20)*100
z=matrix(kriging.mean(Xnew=a,m=model),20,20)*1000

pdf('YWo3_noxavoc_1600_19th.pdf',width=5,height=5)

par(mar=c(4.5,4.5,1,1))

image2D(z=z,x=x,y=y,resfac=8,contour=list(lwd=1.5,labcex=1),
    colkey=F,
        col=jet2.col(n=100),cex.axis=0.8,
        xlab='Anthropogenic VOC emissions (% of base case)',
        ylab='NOx emissions (% of base case)')

abline(v=100,lwd=1.5,lty=2,col='grey50')
abline(h=100,lwd=1.5,lty=2,col='grey50')

dev.off()
```

B.7 R code for BMC and history matching

B.7.1 BMC

```
rm(list=ls())

library(ncdf)
```

```
library(DiceKriging)
library(truncnorm)

hyper<-read.table('./hyper.txt',header=F)

colnames(hyper)<-c('NO2 phot','O3+NO','NO2+O3','O3 phot',
                   'O1D+M','O1D+H2O','NO2+OH','HO2+NO',
                   'OH+CH4','CH2O phot','C2O3+NO','PAR+OH',
                   'UK ISOP','EU ISOP','UK NO','EU NO',
                   'UK NO2','EU NO2','UK PAR','EU PAR',
                   'UK XYL','EU XYL','UK ETH','EU ETH',
                   'UK CO','EU CO','UK OLE','EU OLE',
                   'O3 BC','CO BC','O3 dep V')

output<-0.0
O<-0.097
O.sig<-0.0025
n<-100000
O<-rep(O,n)

for (i in 1:576)
    {
    setwd(paste('./data',i,sep=''))
    nc<-open.ncdf('CCTM_e1a_Linux2_x86_64pg.EU09.ACONC.2006200.
        ncf')
    output[i]<-get.var.ncdf(nc,'O3',c(41,17,1,17),c(1,1,1,1))
    close.ncdf(nc)
    setwd('..')
    }

model<-km(design=hyper,response=output)

kriging.mean<-function(Xnew,m)
    {
     predict.km(m,Xnew,'UK',se.compute=F,checkNames=F)$mean
    }

means<-rep(1,12)
length(means)<-31
SDs<-c(.2,.1,.15,.3,.5,.08,.4,.3,.1,.4,.5,.3)
length(SDs)<-31
```

```
mins<-c(rep(.5,12),rep(.7,19))
maxes<-c(rep(2.0,12),rep(1.3,19))

mins[31]<-.5
maxes[13]<-2.0
maxes[14]<-2.0
maxes[15]<-2.0
maxes[16]<-2.0

dist<-data.frame(means,SDs,mins,maxes)
rownames(dist)<-colnames(hyper)
MCmatrix<-matrix(1,n,31)
colnames(MCmatrix)<-colnames(hyper)

for (i in 1:12)
  {
  MCmatrix[,i]<-rtruncnorm(n,a=dist$mins[i],b=dist$maxes[i],
  mean=dist$means[i],sd=dist$SDs[i])
  }

for (i in 13:31)
  {
  MCmatrix[,i]<-runif(n,min=dist$mins[i],max=dist$maxes[i])
  }

Y<-kriging.mean(MCmatrix,m=model)
y<-kriging.mean(matrix(1,1,length(means)),m=model)

densY<-density(Y)
densY$y<-densY$y/sum(densY$y)

par(mfcol=c(1,1))
par(mar=c(5,5,2,2))
plot(densY,lwd=2,main='',xlim=c(.04,.14),ylim=c(0,.008),
      col='blue',xlab='Ozone conc (ppb)',cex.lab=1.5,
      cex.axis=1.5,xaxt='n')
axis(1,at=c(.04,.06,.08,.1,.12),labels=c('40','60','80',
        '100','120'),cex.axis=1.5)
abline(v=y,lty=2,lwd=3)
abline(v=O[1],lty=2,lwd=3,col='darkseagreen')
legendText=c('Emulated\n output',
```

```
                'Observed\n value')
legend(legend=legendText,col=c('black','darkseagreen'),
       lty=c(2,2),lwd=3,x=0.105,y=0.0075,horiz=F,
       cex=1.3,y.intersp=2,bty='o')

#qqnorm(Y,pch=1)
#qqline(Y,col='red',lwd=2)

k<-(1/(sqrt(2*pi)))/O.sig
diff<-O-Y
diffBySig<-diff/O.sig
like.Y<-k*(exp(-0.5*(diffBySig*diffBySig)))
post.Y<-like.Y/n
C<-sum(post.Y)
post.Y.norm<-post.Y/C

post.mu<-sum(post.Y.norm*Y)
post.sigma<-sqrt(sum((Y-rep(post.mu,n))*(Y-rep(post.mu,n))*post
    .Y.norm))

post.densY<-density(Y,weights=post.Y.norm)
post.densY$y<-post.densY$y/sum(post.densY$y)

pdf(file='HARbmc15.pdf',width=7,height=7)

par(mar=c(5,5,2,2))
plot(densY,lwd=2,main='',xlim=c(.04,.12),ylim=c(0,.018),
     col='blue',xlab='Ozone conc. (ppb)',cex.lab=1.5,
     cex.axis=1.5,xaxt='n')
axis(1,at=c(.02,.04,.06),labels=c('20','40','60'),
       cex.axis=1.5)
lines(post.densY,lwd=2,col='red')
abline(v=y,lty=2,lwd=3)
abline(v=O[1],lty=2,lwd=3,col='darkseagreen')
legendText=c('Emulated\n output',
             'Observed\n value',
             'Uncalibrated\n distribution',
             'Calibrated\n distribution')
legend(legend=legendText,col=c('black','darkseagreen',
       'blue','red'),lty=c(2,2,1,1),lwd=3,x=0.055,
       y=0.017,horiz=F,cex=1.3,y.intersp=2,
```

```
        bty='o')

dev.off()

DATA<-data.frame(MCmatrix,Y,post.Y,post.Y.norm)
DATA<-DATA[order(DATA$post.Y,decreasing=T),]

pdf(file='calibinputs.pdf',width=20,height=14.188)

m<-matrix(c(1,1,2,3,4,5,6,
            1,1,7,8,9,10,11,
            12,13,14,15,16,17,18,
            19,20,21,22,23,24,25,
            26,27,28,29,30,31,32),ncol=7, by=T)
layout(m)

screen(1)
plot(densY,lwd=2,main='',ylim=c(0,.02),xlim=c(0.05,0.15),
     col='blue')
lines(post.densY,lwd=2,col='red')
abline(v=y,lty=2,lwd=2)
abline(v=O[1],lty=2,lwd=2,col='green')

for (i in 1:31)
  {
  densOut<-density(MCmatrix[,i])
  densOut$y<-densOut$y/sum(densOut$y)
  post.densOut<-density(MCmatrix[,i],weights=post.Y.norm)
  post.densOut$y<-post.densOut$y/sum(post.densOut$y)
  screen(i+1)
  plot(post.densOut,lwd=2,main=colnames(hyper)[i],col='red',
  xlab='Scaling Factor')
  lines(densOut,lwd=2,col='blue')
  }

dev.off()
```

B.7.2 History matching

```
rm(list=ls())
```

```
library(ncdf)
library(DiceKriging)
library(truncnorm)

hyper<-read.table('./hyper.txt',header=F)

colnames(hyper)<-c('NO2 phot','O3+NO','NO2+O3','O3 phot',
                   'O1D+M','O1D+H2O','NO2+OH','HO2+NO',
                   'OH+CH4','CH2O phot','C2O3+NO','PAR+OH',
                   'UK ISOP','EU ISOP','UK NO','EU NO',
                   'UK NO2','EU NO2','UK PAR','EU PAR',
                   'UK XYL','EU XYL','UK ETH','EU ETH',
                   'UK CO','EU CO','UK OLE','EU OLE',
                   'O3 BC','CO BC','O3 dep V')

output<-0.0
O<-0.097
O.sig<-0.0025
O.var<-O.sig^2
M.sig<-0.005
M.var<-M.sig^2
n<-100000
O<-rep(O,n)
O.var<-rep(O.var,n)
M.var<-rep(M.var,n)

for (i in 1:576)
    {
    setwd(paste('./data',i,sep=''))
    nc<-open.ncdf('CCTM_e1a_Linux2_x86_64pg.EU09.ACONC.2006200.
        ncf')
    output[i]<-get.var.ncdf(nc,'O3',c(41,17,1,17),c(1,1,1,1))
    close.ncdf(nc)
    setwd('..')
    }

model<-km(design=hyper,response=output)

kriging.mean<-function(Xnew,m)
    {
```

```
    predict.km(m,Xnew, 'UK',se.compute=F,checkNames=F)$mean
   }

kriging<-function(Xnew,m)
   {
    predict.km(m,Xnew, 'UK',se.compute=T,checkNames=F)
   }

mins<-rep(.5,31)
maxes<-rep(2.0,31)

MCmatrix<-matrix(1,n,31)
colnames(MCmatrix)<-colnames(hyper)

for (i in 1:31)
  {
  MCmatrix[,i]<-runif(n,min=mins[i],max=maxes[i])
  }

K<-kriging(MCmatrix,m=model)
Y<-K$mean
S<-K$sd
y<-kriging.mean(matrix(1,1,length(mins)),m=model)

numerator.I<-(O-Y)^2
denominator.I<-(S^2)+M.var+O.var
I.sq<-numerator.I/denominator.I
I<-sqrt(I.sq)

DATA<-data.frame(MCmatrix,Y,S,I)
index<-DATA$I<3
plDATA<-DATA[index,]
plRanges<-apply(plDATA[,1:31],2,range)

index2<-DATA$I>=3
implDATA<-DATA[index2,]

hist(DATA$Y,breaks=brks,col='lightblue',main='')
hist(plDATA$Y,breaks=brks,add=T,col='blue')
```